

---

# **Methodology and Software for Joint Modelling of Time-to-Event Data and Longitudinal Outcomes Across Multiple Studies**

---

Thesis submitted in accordance with the requirements of the University  
of Liverpool for the degree of Doctor in Philosophy by

Maria Elisabet Sudell

February 2018



# Abstract

---

**Thesis Title:** *Methodology and Software for Joint Modelling of Time-to-Event Data and Longitudinal Outcomes Across Multiple Studies*

**Author:** *Maria Sudell*

**Introduction and Aims:** Univariate joint models for longitudinal and time-to-event data simultaneously model one outcome that is repeatedly measured over time, with another outcome which measures the time until the occurrence of an event. They have been increasingly used in the literature to account for dropout in longitudinal studies, to include time-varying covariates in time-to-event analyses, or to investigate links between longitudinal and time-to-event outcomes. Meta-analysis is the quantitative pooling of data from multiple studies. Such analyses can provide increased sample size and so detect small covariate effects. Modelling of multi-study data requires accounting for the clustering of individuals within studies and careful consideration of heterogeneity between studies. Research concerning methodology for modelling of joint longitudinal and time-to-event data in a multi-study or meta-analytic setting does not currently exist. This thesis develops novel methodologies and software for the modelling of multi-study joint longitudinal and time-to-event data.

**Methods:** A review of current reporting standards of analyses applying joint modelling methodology to single study datasets, with a view to future Aggregate Data Meta-Analyses (AD-MA) of joint data is undertaken. Methodology for the one and two-stage Individual Participant Data Meta-Analyses (IPD-MA) are developed. A software package in the R language containing functionalities for various aspects of multi-study joint modelling analyses is built. The methodology and software is implemented in a real hypertension dataset, and also is tested in extensive simulation studies.

**Results:** Reporting of model structure was amongst the areas identified for improvement in the reporting of joint models employed in single study applied analyses. Sufficient information was reported in the majority of studies for them to contribute to future AD-MA. Guidelines developed to ensure good quality two-stage IPD-MA of joint data were presented, designed to ensure only parameters with comparable interpretations are pooled. A range of one-stage models, each accounting for between study heterogeneity in varying ways, were described and applied to real data and simulation analyses. Models employing study level random effects were found unreliable for the investigated association structure, however fixed effect approaches or those that stratified baseline hazard by study were more reliable. The benefit of using joint models over separate time-to-event models in the presence of significant association between the longitudinal and time-to-event outcomes in both one and two-stage analyses was established. Novel software capable of one or two-stage analyses of large multi-study joint datasets was demonstrated in both the real data and simulation analyses.

**Conclusions:** Reporting of joint modelling structure in single study applied analyses should be maintained and improved. Two-stage meta-analyses of joint modelling results should take care to pool only parameters with comparable interpretations. In meta-analyses, investigators should employ a joint modelling approach when association is known or suspected between the longitudinal and time-to-event outcomes. Further work into meta-analytic joint models is required to expand the range of available multi-study joint modelling structures, to allow for multivariate joint data, and to employ multivariate meta-analytic techniques in a two-stage meta-analysis.

# Acknowledgements

---

I would like to thank everyone who has helped to make this thesis possible. Firstly, I would like to thank my supervisors Professor Catrin Tudur Smith and Dr Ruwanthi Kolamunnage-Dona for their excellent guidance and constant support throughout my PhD.

Additionally, I would like to thank Professor François Gueyffier for allowing access to data from the INDANA collaboration, and providing comments concerning the real data applications of the developed methods.

I'd also like to thank everyone in the Department of Biostatistics. I have been fortunate to be based in a department which has both a strong research tradition and a welcoming community. Especially, I would like to thank Dr Laura Bonnett for answering questions about everything from survival analysis to which forms to complete in order to submit a thesis, and Dr Sarah Nevitt for talking to me about meta-analyses, and various thesis formatting issues. In addition, I would like to thank Dr Graeme Hickey for his valuable advice when creating my first R package and for many discussions about joint modelling, and also to Dr David Hughes for discussions about longitudinal and joint analyses, and for teaching me to how not to break condor when performing simulation studies.

Finally, I cannot thank my family and friends enough for supporting and encouraging me during my PhD. Especially I would like to thank my husband Mark, for his unceasing support, for proof-reading my chapters, for keeping me supplied with tea and biscuits, and for too many other things to mention. Thank you.



# Table of Contents

---

Abstract.....	i
Acknowledgements.....	ii
Table of Contents.....	iii
Abbreviations.....	viii
Notation .....	ix
Publications of Work in this Thesis .....	xi
Chapter 1: Introduction .....	1
1.1    Meta-analysis.....	1
1.1.1    Individual Patient Data and Aggregate Data.....	2
1.1.2    Heterogeneity .....	5
1.1.3    Bias .....	5
1.1.4    Models for Meta-Analysis .....	6
1.2    Longitudinal Data .....	7
1.3    Meta-Analysis of Longitudinal Data .....	9
1.3.1    Issues surrounding Longitudinal Data Meta-Analysis.....	9
1.3.2    IPD meta-analysis of longitudinal data .....	10
1.3.3    Aggregate data meta-analysis of longitudinal data .....	13
1.3.4    Comparison of IPD-MA and AD-MA of longitudinal data .....	17
1.3.5    Conclusions .....	18
1.4    Time-to-event data .....	18
1.5    Meta-Analysis of Time-to-Event Data .....	22
1.5.1    Issues surrounding Time-to-Event Data Meta-Analysis.....	22
1.5.2    Individual participant data meta-analysis of time-to-event data .....	22
1.5.3    Aggregate data meta-analysis of time-to-event data .....	25
1.5.4    Comparison of IPD and AD MA of time-to-event data .....	26
1.5.5    Conclusions .....	26
1.6    Joint models for longitudinal and time-to-event data.....	26
1.6.1    General Structure of a Joint Model.....	27
1.6.2    Association Structures .....	28
1.6.3    Currently available software .....	29
1.6.4    Discussion.....	29

1.7	Meta-Analysis of Joint Data .....	30
1.8	Thesis Aims and Structure.....	31
1.8.1	Motivation for research .....	31
1.8.2	Aims of Thesis .....	31
1.8.3	Thesis structure.....	32
Chapter 2: Review of reporting of joint models in the literature .....		34
2.1	Motivation for the review.....	34
2.2	Review Methods .....	36
2.2.1	Identification of Papers.....	36
2.2.2	Data extraction.....	37
2.3	Results.....	39
2.3.1	Characteristics of identified studies.....	39
2.3.2	Did studies report sufficient information to contribute to AD-MA? .....	50
2.4	Discussion.....	52
2.4.1	Recommendations for reporting of future single study joint analyses .....	52
2.4.2	Concluding Remarks.....	55
Chapter 3: Methods for Meta-Analysis of Joint Longitudinal and Time-to-Event Data .....		56
3.1	Description of assumed data type .....	56
3.1.1	General notation .....	56
3.2	Methods for Two-stage IPD Meta-Analyses .....	57
3.2.1	Preliminaries .....	58
3.2.2	Stage 1 – Modelling of joint data within studies .....	61
3.2.3	Stage 2 – Pooling of results between studies .....	62
3.2.4	Recommendations for a two-stage IPD-MA of joint data.....	63
3.3	Methods for One-stage IPD Meta-Analyses.....	66
3.3.1	Preliminaries .....	67
3.3.2	Methods to account for between study heterogeneity .....	68
3.3.3	Number of parameters in one-stage joint MA models.....	83
3.4	Discussion of meta-analytic joint modelling methods.....	84
3.4.1	Discussion of methods for two-stage joint IPD-MA.....	84
3.4.2	Discussion of methods for one-stage joint IPD-MA.....	88
3.4.3	Concluding Remarks.....	90
Chapter 4: R Software package for Meta-Analysis of Joint Longitudinal and Time-To-Event Data.....		93

4.1	Installing the <i>joineRmeta</i> package.....	94
4.2	Data.....	94
4.2.1	Simulated data .....	95
4.2.2	Single study joint model fits.....	95
4.2.3	One-stage fits to multi-study joint data.....	96
4.3	Analysis Functions.....	96
4.3.1	Methods for Two-stage Meta-Analyses.....	96
4.3.2	Methods for One-Stage Meta-Analyses.....	103
4.4	Exploratory Functions .....	118
4.4.1	Data Simulation.....	118
4.4.2	Data Preparation.....	123
4.4.3	Data Visualization.....	127
4.5	Discussion.....	134
Chapter 5: Methodology Supporting the Implementation of the <i>joineRmeta</i> Package ..		135
5.1	Analysis Functions.....	135
5.1.1	Methods for Two-stage Meta-Analyses.....	135
5.1.2	Methods for One-Stage Meta-Analyses.....	137
5.2	Simulation of multi-study joint data .....	161
5.3	Discussion.....	166
Chapter 6: Real Data Applications of Joint Meta-Analytic Methods .....		167
6.1	Description of the INDANA dataset .....	167
6.1.1	Comparison of demographics between studies .....	168
6.1.2	Longitudinal Data .....	174
6.1.3	Time-to-event data .....	183
6.2	Two-stage Meta-Analysis of INDANA dataset.....	188
6.2.1	Methods.....	188
6.2.2	Results.....	191
6.3	One-stage Meta-Analysis of INDANA dataset.....	197
6.3.1	Methods.....	197
6.3.2	Results.....	199
6.4	Discussion of Joint Meta-Analysis of Real Data .....	216
6.4.1	Discussion of Two-Stage joint MA of INDANA data .....	216
6.4.2	Discussion of One-Stage joint MA of INDANA data .....	217
6.4.3	Comparison of one and two-stage approaches .....	221

6.4.4	Limitations of conducted analyses.....	221
6.4.5	Concluding remarks .....	222
Chapter 7: Simulation investigation of joint meta-analytic methods.....		223
7.1	Simulations investigating two-stage methods.....	223
7.1.1	Simulation of joint data for two-stage simulations .....	224
7.1.2	Two-stage models fitted to simulated joint data.....	227
7.1.3	Two-stage simulation study results .....	228
7.2	Simulations investigating one-stage methods.....	237
7.2.1	Simulation of joint data for one-stage simulations.....	237
7.2.2	One-stage models fitted to simulated joint data.....	242
7.2.3	One-stage simulation study results .....	245
7.3	Discussion of simulation studies .....	264
7.3.1	Discussion of the Two-Stage Simulation Investigation .....	264
7.3.2	Discussion of the One-Stage Simulation Investigation .....	265
7.3.3	Comparison of one and two-stage simulations .....	266
7.3.4	Concluding Remarks.....	270
Chapter 8: Conclusions and Further Work.....		272
8.1	Overview of Thesis and Key Messages .....	272
8.2	Implications for research .....	274
8.3	Planned future work .....	276
8.3.1	Planned future methodological research .....	276
8.3.2	Planned developments to the <i>joinerMeta</i> package.....	278
8.3.3	Planned future applied investigations .....	281
8.4	Conclusion.....	281
Bibliography .....		282
Appendices.....		296
Appendix 1: Search Strategies used in Review of Current Reporting Standards of Joint Models in the Literature .....		296
Database : Scopus .....		296
Database : PubMed.....		296
Database: Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) <1946 to Present with Daily Update>.....		296
Appendix 2: Additional longitudinal SBP trajectory plots for INDANA dataset .....		298
SBP and time to Death .....		298

SBP and time to MI .....	299
SBP and time to stroke.....	300
Appendix 3: Event Rates for Two-Stage Simulation Studies.....	301
Appendix 4: One stage simulation graphs showing point estimates and confidence intervals.....	302
Simulation Set 1: Varying levels of association.....	302
Simulation Set 2: Varying number of included studies.....	314
Simulation Set 3: Varying levels of between study heterogeneity .....	318
Appendix 5: Association Structure Estimates for One-Stage Analyses of INDANA data .	322

## Abbreviations

ABBREVIATION	EXPANSION
AD	Aggregate Data
AD-MA	Aggregate Data Meta-Analysis
AFT	Accelerated Failure Time
AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
CVD	Cardiovascular disease
EM ALGORITHM	Expectation-Maximisation algorithm
HR	Hazard Ratio
IID	Independently and Identically distributed
IPD	Individual Participant Data
IPD-MA	Individual Participant Data Meta-Analysis
LQ	Lower Quartile
MA	Meta-Analysis / meta-analyses / meta-analytic
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MI	Myocardial Infarction
MLE	Maximum Likelihood Estimate
PH	Proportional Hazards
SD	Standard Deviation
SE	Standard Error
UQ	Upper Quartile

## Notation

Note that parameters displayed in bold could indicate a vector containing multiple elements e.g.  $\alpha = (\alpha^{(2)}, \alpha^{(3)})$ , or a matrix e.g.  $A$ .

TERM	DESCRIPTION
$A$	Covariance matrix for study level (level 3) random effects
$a$	Abscissa for Gauss-Hermite quadrature generated by generic algorithm
$a_{(3)k}$	Abscissa for Pseudo-adaptive Gauss-Hermite quadrature for study level random effects
$a_{(2)ki}$	Abscissa for Pseudo-adaptive Gauss-Hermite quadrature for individual level random effects
$\alpha$	Collection of association parameters
$\alpha^{(2)}$	Association parameter linked to individual level (level 2) random effects
$\alpha^{(3)}$	Association parameter linked to study level (level 3) random effects
$B$	Choleski decomposition of negative second derivative of log likelihood
$\beta$	Collection of fixed effect coefficients
$\beta_1$	Fixed effect coefficients, with first element of subscript identifying as from longitudinal sub-model
$\beta_2$	Fixed effect coefficients, with first element of subscript identifying as from time-to-event sub-model
$b$	Zero mean normally distributed random effects
$b_{ki}^{(2)}$	Individual level (level 2) random effects
$b_k^{(3)}$	Study level (level 3) random effects
$D$	Covariance matrix for individual level (level 2) random effects
$\Delta_{ki}$	Censoring indicator for individual $i$ in study $k$
$\varepsilon_{kij}$	Error term from longitudinal sub-model
$f$	Used to denote a function, e.g. a component of the likelihood
$g$	Generic counter used to count through e.g. the events when calculating the baseline hazard or through the score vector when updating the fixed effects from a sub-model
$h$	Generic counter used to count through e.g. the events when calculating the baseline hazard or through the score vector when updating the fixed effects from a sub-model
$I$	Information matrix
$i$	Character used to identify individuals within studies
$j$	Character used to identify measurements within individuals within studies
$k$	Character used to identify studies
$K$	Total number of studies included in meta-analysis
$\kappa$	Scale parameter for Gompertz distribution used during simulation studies
$L(\Omega)$	Likelihood based on complete data
$\ell(\Omega)$	Log-likelihood based on complete data
$\lambda_{ki}(t)$	Hazard function for individual $i$ from study $k$ based on time $t$
$\lambda_0(t)$	Baseline hazard function common across studies
$\lambda_{0k}(t)$	Baseline hazard function stratified by study
$M$	Used to represent a particular model in Section 3.3.3, calculation of the number of parameters present in a one-stage joint model.

$m_{ki}$	The number of longitudinal measurements recorded for individual $i$ in study $k$
$n_k$	The number of individuals included in study $k$
$P$	The degrees of freedom of the $\chi^2$ distribution used when comparing two nested models
$p_1$	The number of fixed effects included in the longitudinal sub-model
$p_2$	The number of fixed effects included in the time-to-event sub-model
$q$	The number of individual level random effects
$r$	The number of study level random effects
$\varrho$	The number of association parameters to be estimated
$S$	Score vector
$\sigma_e^2$	Variance for longitudinal error term $\varepsilon_{kij}$
$\sigma_A^2$	Variance of study level random effect if only a single study level random effect is included in the one stage meta-analytic joint model
$t_{kij}$	Longitudinal time variable for individual $i$ in study $k$ at time point $j$
$\tau^2$	Term quantifying between study heterogeneity
$T_{Cki}$	Censoring time for individual $i$ in study $k$
$T_{Eki}$	Event time for individual $i$ in study $k$
$T_{Ski}$	Survival time (minimum of event and censoring time) for individual $i$ in study $k$
$v_k$	Variance of treatment effect from study $k$ used in two stage MA
$v$	Iteration counter used in Newton-Raphson estimation of $\beta_2$ parameters in M-step
$\omega$	Shape parameter for Gompertz distribution used during simulation studies
$w_k$	Weight for study $k$ used in two stage MA
$W_{1ki}(t)$	Function of longitudinal sub-model involved in the association structure
$W_{2ki}(t)$	Term present in time-to-event sub-model as part of association structure
$X_1$	Design matrix for covariates included with fixed effects in the longitudinal sub-model
$X_2$	Design matrix for covariates included with fixed effects in the time-to-event sub-model
$Y_{kij}$	Longitudinal outcomes recorded for individual $i$ in study $k$ at time point $j$
$Z_{ki}^{(2)}$	Design matrix for covariates assigned individual level (level 2) random effects
$Z_k^{(3)}$	Design matrix for covariates assigned study level (level 3) random effects
$\Omega$	Complete data, used in description of full likelihood in one stage model



## Publications of Work in this Thesis

---

Chapter 2 has been published in:

*Sudell, M., Kolamunnage-Dona, R., & Tudur-Smith, C. (2016). Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis. BMC Medical Research Methodology, 16, 168. <http://doi.org/10.1186/s12874-016-0272-6>*

Work on the two stage MA of joint longitudinal data presented in Chapters 3, 6 and 7 has been published in:

*Sudell M, Tudur Smith C, Gueyffier F, Kolamunnage-Dona R (2017). Investigation of 2-stage meta-analysis methods for joint longitudinal and time-to-event data through simulation and real data application. Statistics in Medicine.1–18. <https://doi.org/10.1002/sim.7585>*



# Chapter 1 : Introduction

---

This thesis aims to investigate the meta-analysis of joint longitudinal and time-to-event data. Joint models allow simultaneous assessment of longitudinal and time-to-event outcomes, and are useful in cases of informative dropout, or time-varying covariates for time-to-event data, or where a relationship is suspected between longitudinal and time-to-event outcomes. Meta-analyses allow all available evidence from multiple studies to be used to answer a research question, whilst accounting for any between study variability. Linking these two areas would ensure that multi-study longitudinal and time-to-event data can be appropriately analysed, however currently methodology and software to permit such analyses to be conducted is lacking. This thesis will combine these areas, providing methodology and software for use in the meta-analysis of joint data.

To accomplish this, the thesis will firstly assess the current reporting standards of joint models applied to single studies, with a view to pooling published results in future aggregate data meta-analyses. Following this, frequentist methods for both two and one-stage meta-analysis of individual participant joint data will be proposed, and assessed through application to both real and simulated datasets. A novel software package developed to facilitate the application of the proposed methods is presented and discussed. Finally areas where future research would be beneficial are examined.

In this chapter an introduction to the concepts of meta-analysis and joint modelling is given, in order to introduce the terminology and methodology that this thesis aims to extend. A review of current methodology for the meta-analysis of longitudinal or time-to-event data is also presented. The chapter concludes with a statement of the aims and structure of the thesis.

Throughout the thesis, reference to computer packages, functions or code is identified through a change in font e.g. the R package developed during this research is identified in the text by *joineRmeta*.

## 1.1 Meta-analysis

One of the main aims of medical research is patient wellbeing, both protecting patients (whether participating in clinical trials or undergoing general treatment) from potentially harmful or less effective treatments, but also identifying treatments that improve standard care (either by being more effective, or by causing fewer side effects). A common practice

is to undertake a systematic review, possibly combined with a meta-analysis, to assess what investigations have already been conducted in an area, and to compile the available information in order to inform a clinical decision concerning a treatment, or to demonstrate the need for further clinical trials.

A systematic review is a structured assessment of studies relating to the research question of interest (for example the benefits of a given treatment over the current standard for a specified disease and population). This process identifies published (and potentially unpublished) studies that address this research question, and collates their results into one report, thereby providing clinicians with a concise summary of the current information concerning a treatment.

A meta-analysis (MA) is similar to a systematic review in that it collates available evidence on a subject, but it quantitatively collates this evidence rather than qualitatively evaluating it. Glass [1] defined a meta-analysis to be “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings”. The estimates of treatment effect sizes from different studies, for example, are pooled. An overall summary treatment effect is then calculated, based on the information from the available studies. Therefore whilst a systematic review presents the available evidence, a meta-analysis generates a numeric quantification of the available evidence, giving a stronger basis for clinical decisions [2].

Section 1.1 gives a general background to the area of meta-analyses, with details of meta-analysis for longitudinal or time-to-event data given in Sections 1.3 and 1.5 respectively. Further information concerning meta-analytic methods is available in Whitehead [2].

### **1.1.1 Individual Patient Data and Aggregate Data**

A meta-analysis can be based on Aggregate Data (AD), Individual Participant or Patient Data (IPD) or a mix of the two (see Riley et al [3], Sutton et al [4] and Donegan et al [5]). Some studies have noted the increased power in an IPD regression analysis compared to an AD regression analysis [6]. These two data types will now be defined and discussed.

#### ***1.1.1.1 Aggregate Data Meta-Analysis***

An Aggregate Data Meta-Analysis (AD-MA) can be performed using study level summary statistics provided in study reports, or by study authors through direct communication. Information is not available for each individual, only for each study, for example the exact

characteristics of each study participant would not be known, however the proportions of the study population falling into each group of a categorical variable may be known.

One advantage of an AD-MA is that it can be completed with less resource expenditure than an IPD-MA, as there is no lengthy data collection process. Additionally an AD-MA does not require as high a level of statistical knowledge as an IPD-MA, in which data specific methods such as time-to-event analysis may be required. Straightforward computer software is available to aid researchers to easily complete AD-MA without a high level of statistical programming, for example the RevMan software provided by the Cochrane group [7].

A major drawback of AD-MA is that summary statistics may not be uniformly reported over studies. Also some studies may report unadjusted estimates, whilst others report only adjusted, or a mix of the two. There may be differences across the scales used to report results (for example differing depression scales between studies), or the measures used (one study reporting risk differences, another reporting odds ratios), or the scales used (log versus unlogged results [8]). Sufficient information to complete the meta-analysis may not always be included, for example *p* values may be reported in place of odds ratios, or sample sizes may not be stated in the text of the report. This may result in clarification of data being required from study authors, which may extend the time taken to complete the meta-analysis.

Another limitation of AD-MA is that inference concerning patient level covariate effects (for example effect of age of the patient on treatment effect) cannot be made. Only the effect of study level covariates (such as median patient age in each study) can be examined. Any attempt to interpret study level covariates at the patient level can result in ecological bias [9].

Additionally if interest lies in a particular subgroup of patients, for example those over a certain age or with a particular characteristic, only studies reporting the relevant subgroup summary statistics can contribute to the analyses (unless study authors can provide additional information). Consequently subgroup analyses can be hard to perform, or may only contain a subset of the identified studies. Furthermore an investigation into the reasons for any heterogeneity between included studies may be hampered by having access only to aggregate data. AD-MA can account for heterogeneity by using a random effects model [2, 10] or study level meta-regression [11], but cannot fully investigate the

causes of the heterogeneity as only study level covariates can be used in an attempt to explain the presence of heterogeneity.

Some of these problems can be overcome, through additional communication with the study authors, or standardisation of scales or measures used. However, sometimes sufficient information to complete the required analyses cannot be obtained, and so the researcher has to choose between a restricted analysis or the possibly lengthy process of obtaining IPD.

#### *1.1.1.2 Individual Participant Data Meta-Analysis*

Individual Participant Data Meta-Analyses (IPD-MA) are considered the gold standard of meta-analysis [12]. Stewart and Parmar [13] suggest that if possible an IPD meta-analysis should be performed in preference to an AD meta-analysis, as it reduces bias and allows further investigation into additional research questions (for example differing treatment effects between subgroups). Additionally IPD-MA allow investigations into the reasons behind heterogeneity between studies, including examination of interaction effects between covariates. Furthermore, having access to the original data allows more in-depth data checking, and replication of the results given in the study report, which gives more credibility to the data and so the final output. Finally, obtaining IPD allows those designing the meta-analysis to have full choice of methods used. If only aggregate data is obtained, the meta-analysis investigators cannot change the methods used to obtain the study level results.

However, seeking IPD for a meta-analysis is often time consuming and resource intensive. For example, older datasets may no longer be available or may only exist in hard copy format, or there may be issues sharing the data. Nevitt et al [14] provide a review of retrieval of IPD for systematic reviews. In addition to issues obtaining data, IPD-MA may require more complex data analysis methods (for example for longitudinal or time-to-event data), resulting in a higher necessary level of statistical expertise, and knowledge of statistical software in the team conducting the MA [15]. A discussion of the benefits and problems of seeking IPD is given in Chapter 18 of the Cochrane handbook [16], and the benefits of an IPD analysis over an AD analysis are discussed by Stewart [17]. Additionally Tudur Smith et al [18] compared the results of IPD and AD meta-analyses, concluding that IPD-MA allow additional investigations to be conducted, but as the results are often similar it may be beneficial to conduct an AD-MA initially, and assess what benefit would be gained by conducting an IPD-MA.

#### 1.1.1.3 *IPD and AD during this research*

Methods developed during this research assume availability of IPD from each study included in the meta-analysis. This is due to the fact that joint models are complex, and so IPD is necessary to ensure that the model structure is correctly estimated. In any case where the methodology presented has applications to AD-MA, these are highlighted and discussed.

#### 1.1.2 Heterogeneity

An important issue in meta-analysis is heterogeneity. Heterogeneity is defined to be any variability between studies included in the meta-analysis [19], and can be divided into clinical heterogeneity (differences in study populations, interventions and outcomes), and methodological or statistical heterogeneity (differences in risks of bias between studies and differences in study design). Heterogeneity is quantified using various statistics including the  $I^2$  statistics and Cochrane's  $Q$  (see Higgins and Thompson [20] for more detail). It is important to account for heterogeneity in a meta-analysis, as otherwise conclusions drawn could be misleading. If there is no evidence of heterogeneity between included studies, data from the studies can be pooled without difficulty. However, if there is heterogeneity between studies, this needs to be accounted for in the analysis, for example by using a meta-regression (Thompson et al [11]) or random effects model (DerSimonian and Laird [10]). If there is a large amount of heterogeneity between studies (for example significant estimates of treatment effect in opposite directions reported by included studies), it may not be appropriate to pool results from different studies, and so only a narrative summary of the results should be given. A benefit of using IPD over AD is that any heterogeneity present can be investigated using variables in the data.

#### 1.1.3 Bias

Bias is defined as some systematic error [21] that can cause the results of a study to be under or overestimated. Bias is not the same as random error, as bias is a systematic deviation from the true value within a study rather than a varying deviation. Generally meta-analyses assess the risk of bias for included studies rather than attempting to quantify the amount of bias. Including studies at more risk of bias could lead to more erroneous conclusions being drawn: a meta-analysis of more rigorous studies which are at less risk of bias is generally more reliable.

As bias is not a focus of this thesis, it will not be discussed further. More details concerning bias can be found in Whitehead [2] and in Chapter 8 of the Cochrane Handbook [21].

#### 1.1.4 Models for Meta-Analysis

In the coming section, models commonly encountered in meta-analysis are discussed, including one-stage and two-stage models, as well as fixed and random effects meta-analyses. Models are discussed in general terms, with meta-analytic models specific to joint longitudinal and time-to-event data presented in Chapter 3.

##### 1.1.4.1 *One-stage models*

A one-stage model is where data from different studies are held in one large meta-dataset, and a single model is fitted to the dataset. The study that the data originated from is identified in some way in the model, such as by stratifying by study, or including a fixed or random effect study indicator in the model. Therefore, the model fitted is a hierarchical model, with individuals in the dataset nested within the included studies. To fit a one-stage model, IPD must be obtained from the included studies. One-stage MA models for joint data are discussed further in Section 3.3.

##### 1.1.4.2 *Two-stage models*

A two-stage model is where separate models are fitted to data from each study, and then the study level results from these separate models are pooled using general meta-analytic methods. As with a one-stage model, a two-stage model requires the IPD to be obtained from the included studies. However, the second stage of a two-stage model, where the study level treatment estimates are pooled, is comparable to performing a meta-analysis on aggregate data (AD-MA). Two-stage joint MA models are discussed further in Section 3.2.

###### 1.1.4.2.1 *Fixed effect and random effects models*

In the second stage of a two-stage model, or an AD-MA, fixed effect meta-analyses assume that the true treatment effect is constant across all included studies, and that any variation can be attributed to study specific measurement error. These models are straightforward to implement and most meta-analytic software provide functions to fit them. However, as the treatment effect estimate is based solely on the included trials, it may not be generalizable to a wider population. Also, as the treatment effect is considered to be the same across studies, fixed effect models do not account for possible heterogeneity between studies.

Random effects MA assume that the treatment effect in each included study is a realisation from a common random variable [10]. If there is no heterogeneity between the included studies, the fixed effect and random effect models should produce the same results.



Random effects models can account for heterogeneity between studies and so are often used if the included studies appear heterogeneous in nature. Their use is disputed as a random effects MA assumes the included studies are a random sample of all studies that address the research question. However, a meta-analysis expects that all available studies are included in the meta-analysis. In addition, a random effects MA assign more weight to smaller studies than a fixed effect MA when heterogeneity is present, which can be an invalid assumption. Further discussion of fixed versus random effects MA is given in Whitehead [2].

#### 1.1.4.3 Extensions to basic meta-analysis

For more complex data types (including longitudinal data, time-to-event data or multiple correlated outcomes) more advanced methods are needed, because these data types inherently have additional structure that needs to be accounted for in the models. For example, in longitudinal data, measurements recorded from the same individual are correlated. An overview of current methodology for longitudinal and time-to-event MA are given in Sections 1.3 and 1.5. Many papers have been published dealing with data types other than longitudinal or time-to-event, such as cross-sectional data. These include methods for ordinal data [22] for continuous outcomes [23], and methods to combine parallel and cross over trials [24-26].

## 1.2 Longitudinal Data

Longitudinal data (sometimes referred to as repeated measures data) is data repeatedly collected from study participants over time, and unlike crossover trials, study participants generally undertake only one treatment regimen during the study period. By collecting longitudinal data, researchers can investigate differences between groups with particular demographics at a given time, as well as trends over time. Additionally, collection of longitudinal data can be more cost effective than other data types, as it is cheaper to repeatedly measure an individual already included in the study than expending resources in finding and recruiting new individuals [27].

Longitudinal data can be balanced or unbalanced. Balanced data results from studies that plan to measure individuals a predefined number of times, for example a study that takes measurements at baseline, 3 months, 6 months and 1 year. If a set number of measurements was planned, but some individuals have missed measurements, the data is referred to as balanced with missing data. Unbalanced data refers to a dataset where the

number of measurements and the timing of measurements between individuals can vary considerably, and generally there are not specified measurement times. Some methods to analyse longitudinal data rely on having balanced data (for example repeated measures ANOVA), whereas others can incorporate unbalanced data (for example multilevel models, including random effects models), see Diggle et al [28] for further information.

Variables in longitudinal data can be thought of as constant (such as gender), or variable (such as blood pressure) across the study period, denoted time stationary and time varying covariates respectively. Time varying covariates can be split into two groups: Endogeneous and exogeneous (also referred to as internal and external covariates). These are further discussed in Section 1.6.

It is important when modelling longitudinal data to ensure that the structure of the data is correctly accounted for. As measures are repeatedly recorded from each individual, within individual measurements will be more similar than measurements between different individuals. Individual specific random effects can model the correlation between measurements within individuals, as such the mixed effect model accounts for the nested structure of the data.

Various models currently exist for longitudinal data, such as growth curves, Generalised Estimating Equations and mixed effect models. Mixed effect models were described by Laird and Ware [29], and have structure as displayed in equation (1).

$$Y_i = X_i\beta_1 + Z_i b_i + \varepsilon_i \quad (1)$$

In the above equation, the longitudinal measurements recorded for individual  $i$  are represented by  $Y_i$ . This model type contains both fixed and random effects, which are separated out in the model. Fixed effects (represented by  $\beta_1$  terms) are parameters that have a constant value across the study population, and so are sometimes termed population effects. Covariates assigned fixed effects are held in design matrix  $X_i$ . Random effects (represented by  $b_i$  terms) are random variables that follow a zero mean distribution that act on a particular level of the data, e.g. individuals, and quantify how a unit at that level differs from the population average. Covariates assigned random effects are held in design matrix  $Z_i$ , which is generally a subset of the  $X_i$  matrix. The final term  $\varepsilon_i$  is a vector of error terms for each time point.

Mixed effects models allow population effects and trends to be modelled, but can also estimate through random effects how much individuals differ from these estimates. The

exact specification of models used for longitudinal data are discussed where appropriate in later chapters.

Other models for analysis of longitudinal data exist, for example Generalised Estimating Equations (GEE), or growth curves. However, in this thesis only mixed models will be considered, as mixed models are commonly used to represent the longitudinal component in joint model formulation discussed below. The reader is referred to Weiss [27] or Diggle et al [28] for a comprehensive introduction to longitudinal data analysis.

An issue often present in longitudinal datasets is dropout, which can occur in time-to-event studies or in longitudinal studies. Dropout is when an individual provides measurements up to a certain time point, and then leaves the study and so provides no further information. Dropout can occur for reasons unrelated to the study, in which case it is referred to as non-informative dropout, or for reasons of interest to the study, referred to as informative dropout. An example of informative dropout discussed by Henderson et al [30] in a study concerning schizophrenia patients that measured a PANSS (Positive And Negative Symptom Scale) score over time, where researchers were concerned that patients with higher (worse) PANSS scores were less likely to complete the study (i.e. dropout for reasons of interest to the study outcomes before the end of the study period). Informative dropout is one of the motivations for using the joint modelling structure, introduced in Section 1.6.

### **1.3 Meta-Analysis of Longitudinal Data**

The benefits of longitudinal data, such as modelling trends and cost effectiveness, have naturally prompted researchers to pool longitudinal data or the results from longitudinal model fits in meta-analyses. This section describes some of the issues often encountered during a longitudinal MA, followed by a description of current methodology for AD-MA and IPD-MA of longitudinal data.

#### **1.3.1 Issues surrounding Longitudinal Data Meta-Analysis**

As longitudinal data consists of multiple measurements recorded from the same individual, measurements on the same individual will be inherently correlated. As with single trial longitudinal analyses, meta-analytic methods (AD or IPD) that ignore this correlation may be considered inappropriate. When Jones et al [31] examined the methods commonly used in meta-analyses of longitudinal studies, they identified that there seemed to be no universally adopted approach to longitudinal meta-analysis, and that the most common

approach was to perform multiple independent meta-analyses at various time points, thus ignoring the correlation between time points.

Another difficulty in longitudinal meta-analyses is that the time at which measurements are taken may differ considerably between studies. For example, one study may assess individuals only at pre-specified follow-up meetings, whereas others could additionally record data from unscheduled GP appointments. In addition, some studies may record observations over a much shorter time period than other studies, resulting in different maximum follow-up times between included studies. Therefore if separate independent meta-analyses are performed at specific time points, the studies contributing to each meta-analysis is likely to vary, as studies can only contribute to a meta-analysis if they have measured patients at that particular time point. This may lead to a form of publication bias [32].

In addition to these longitudinal specific issues, longitudinal meta-analyses also fall prey to the more common meta-analysis problems, for example differences in sample populations, study design and study measures. These issues may all introduce bias into the meta-analysis. Additional issues may arise if the longitudinal data stems from observational studies rather than randomised control trials, however this thesis assumes data from randomised controlled trials. Approaches for dealing with bias arising from longitudinal observational studies in a meta-analysis are given in Thompson et al [33].

### 1.3.2 IPD meta-analysis of longitudinal data

The Cochrane Handbook [19], Section 9.3.4 “Repeated observations on participants”, suggests that if longitudinal data is to be meta-analysed using IPD, a suitable model that uses the entire follow-up should be fitted (they suggest a time-to-event analysis), or a single summary measure be calculated for each individual that utilises all the available data (for example one value that represents a trend over time of the longitudinal measure). However, it does not give further guidance concerning what a suitable model for longitudinal IPD would be.

Jones et al [31] propose both one and two-stage models for longitudinal IPD-MA (discussed in the following text). They provide methods for when time is considered continuous (e.g. number of days from randomisation) or as a factor (visit number 1, 2...). They assume an overall set of all possible measurement times (where a measurement time is included if it has occurred in any of the studies to be included in the meta-analysis) on the

understanding that any given study will have only recorded a subset of this set of possible measurement times.

### One-stage approaches

Methods presented in the literature that consider time as a factor (e.g. visit number) are firstly discussed, starting with the methods presented by Jones et al [31]. In their model,  $j$  counts longitudinal measurements (with  $j = 1, \dots, m$ , where  $m$  is the maximum possible number of measurements),  $h$  the treatment indicator (for example,  $h = 0, \dots, g - 1$  where  $g$  is the number of treatment groups),  $k$  be the study indicator (for  $k = 1, \dots, K$ ), and  $i$  be the patient indicator (for  $i = 1, \dots, n_{hk}$  where  $n_{hk}$  is the number of individuals in the  $h$ th treatment group in the  $k$ th study). Also, let  $\varepsilon_{hki}$  be the residual error term and  $y_{hki}$  the longitudinal response for individual  $i$  in treatment group  $h$  in study  $k$  at the  $j$ th measurement. Note that Jones et al allowed for multiple separate treatments, but this thesis concerns randomisation only to a control ( $h = 0$ ) or an experimental ( $h = 1$ ) treatment. Then the one-stage model is given by equation (2).

$$y_{hki} = \alpha_{kj} + \beta_{hj} + \varepsilon_{hki} \quad (2)$$

In this equation, the  $\alpha_{kj}$  is the effect of the control treatment at measurement  $j$  in study  $k$ , and  $\beta_{hj}$  is the additional effect of the experimental treatment  $h$  compared to the control treatment at measurement  $j$ . If baseline measurements are to be included ( $j = 0$ ) the  $\beta_{h0}$  term would be restricted to equal zero as there is expected to be no difference in treatment groups at baseline.

Gurrin and Turkovic [34] mention a similar model to Jones et al for meta-analysing studies with factor time points. They discuss a regression model with study specific intercept, and explanatory variable coefficients stratified by trial for non-longitudinal data, and then briefly show how the model can be extended to allow for repeated measures data, by including a participant specific random effect (random intercept) into the model.

If time is a continuous variable (e.g. days since randomisation) the proposed formulation of the one-stage model in Jones et al [31] changes to that shown in equation (3):

$$y_{hki} = \lambda_k + v_h + \alpha_k t_j + \beta_h t_j + \varepsilon_{hki} \quad (3)$$

In this model,  $\lambda_k$  is the intercept term for the control treatment, and  $v_h$  is the difference in intercepts between the experimental treatment  $h$  and control treatment. Additionally,  $\alpha_k$  is the study specific slope over time for the control treatment, and  $\beta_h$  is the difference in

slope between the experimental treatment  $h$  and the control treatment. As before,  $y_{hki}$  represents the longitudinal response and  $\varepsilon_{hki}$  the residual error of the  $i$ th individual in study  $k$  assigned to treatment  $h$  at the  $t_j$  time point. As before this formulation assumes no baseline measurements – if baseline measurements were to be included then the  $v_h$  term would be removed from the model, as the intercepts would then be calculated at baseline, where they are not expected to differ between treatment groups.

In these models, each study is allowed to exhibit a study specific treatment effect over time. In the discrete time model this is achieved through the study specific  $\alpha_{kj}$  term, which can take a different value for each included study at each time point. In the continuous time model, this is achieved through the study specific intercept term  $\lambda_k$  and the study specific slope term  $\alpha_k$ . These terms allow each included study to have differing initial measurements at the earliest time point, and each included study to display differing treatment effects over time. However, across all studies, these models assume that the difference between the effect of the control and experimental treatment  $h$  is fixed across the included studies (neither the  $\beta_{hj}$  term in the discrete time model, nor the  $\beta_h$  term in the continuous time model contain the  $k$  subscript, indicating that they remain constant across included studies).

### 1.3.2.1 Two-stage approaches

Two-stage methods for longitudinal MA may be more straightforward than one-stage methods. Jones et al [31] describe fitting a separate longitudinal model to each included study and then combining the relevant study level treatment estimates using a multivariate meta-analysis framework [35-41].

If time is supplied as a factor, they suggest that the model in equation (2) would be fitted to the data from each study  $k$  separately and an estimate of the value of the mean difference between treatments  $h$  (for  $h = 0, \dots, g - 1$ ) at each time point  $j$  (for  $j = 1, \dots, m$ ) in that study are calculated, denoted by  $\hat{\beta}_{hkj}$ . The values of  $\hat{\beta}_{hkj}$  from each study  $k$  for  $k = 1, \dots, K$  are then combined using a multivariate meta-analysis framework.

If time has been supplied as a continuous variable, Jones et al [31] suggest that the model in equation (3) is fitted to each included study  $k$  separately, and the estimates of the differences in intercepts between the treatment groups ( $\hat{v}_{hk}$ ), and the differences in slopes between the treatment groups ( $\hat{\beta}_{hk}$ ) are extracted. The study level estimates are then combined again using a multivariate meta-analysis framework

For both time as a factor and time as a continuous variable, the study level estimates are weighted using a function of the variance of the estimates themselves, and the covariance with the other estimates stemming from the same study. A more detailed description is given by Jones et al [31].

#### 1.3.2.1.1 Limitations for longitudinal IPD-MA

For any IPD analysis, the issue of additional resource expenditure compared to an AD analysis exists. In addition the research team must have sufficient statistical knowledge to handle the data appropriately. This issue is exacerbated for longitudinal data, when the nested structure of the data (measurements within individuals) must be accounted for. Correctly modelling the data structure is necessary to reduce bias in the analysis, as such, a longitudinal IPD-MA requires even greater statistical knowledge than a normal IPD-MA.

In addition, modelling of trend over time provides an additional consideration. Jones et al [31] note that their model assumes a linear treatment effect over time, which is potentially not suitable for all datasets. To suitably model the trend over time in longitudinal data, complex methods such as splines may be required, again requiring a high level of statistical expertise.

Furthermore the methods discussed by Jones et al [31] here include only fixed treatment effects across studies. They note the issues involved in introducing random effects into longitudinal IPD-MA models, namely that assumptions need to be made whether random effects should be time specific, and whether there should be between study correlation parameters (random correlation parameters). They note the need for future research in this area. Ishak et al [32] discussed random effect models for AD-MA of longitudinal data (see Section 1.3.3) but experienced issues estimating the relevant correlation parameters, potentially due to the relative shortage of reporting of such parameters in study reports.

### 1.3.3 Aggregate data meta-analysis of longitudinal data

Whilst AD-MA methods are not the main focus of the thesis, in the following section a brief summary is given, highlighting some key references.

Longitudinal AD-MA can be divided into two main groups: namely meta-analyses where the treatment effect is allowed to vary over time, and meta-analyses with a constant treatment effect. Ishak et al [32] noted that for aggregate data cases, if each included study contributes one overall estimate of treatment effect to the meta-analysis (i.e. the treatment effect does not vary over time, for example the slopes of fitted regression

models), these estimates can be combined using basic meta-analytic methods, as correlation between time points in the data should have already been accounted for. The report continues by identifying that if a treatment effect is reported at different time intervals within a study, standard methods cannot be used as they do not account for the correlation between the estimates at different time points. Contrary to this fact, many AD-MA that involve longitudinal studies currently do not account for the correlation between the treatment effects reported at different time points [31].

The Cochrane Handbook [19], in Section 9.3.4 “Repeated observations on participants”, suggests that if aggregate longitudinal data is to be meta-analysed, one of the following three methods should be employed (i) Several outcomes should be defined for different follow-ups, for example a shorter term and a longer term analysis be performed, (ii) one time point should be selected at which to perform a meta-analysis (for either a clinical reason or to maximise the amount of data used), (iii) use the longest follow-up data available from each study (i.e. meta-analyse data from the final reported time-point). However, various issues exist with these methods, for example selecting one time point could increase the possibility of reporting bias (as if a study does not report treatment effect at that time point, it is excluded from analysis), or selecting the longest follow-up could lead to increased heterogeneity (as included studies may vary considerably in maximum follow-up time) [19]. Many of these methods, whilst simple to follow, do not use all the available information and may produce biased results [32].

Ishak et al [32] suggest using mixed effect models, where the responses are the treatment effects from each study at each time point, i.e. a mixed effect model, where measurements from each time point are nested within studies. The basic formulation of this model is given in equation (4). It is assumed that there are  $K$  studies in the meta-analysis, with  $m$  possible measurement times across all studies. In the equation,  $Y_k$  is the vector of length  $m$  of the treatment effect estimates from study  $k$  (so  $Y_k$  is a vector of study level summaries of information at each time point from  $t_1$  to  $t_m$ , without outcome missing where a timepoint was not recorded in the study).  $X_k$  is the  $m \times p$  design matrix for the fixed effects for study  $k$ ,  $Z_k$  is the  $m \times r$  design matrix for the random effects for study  $k$ ,  $\beta$  is the vector of length  $p$  of fixed effects,  $b_k$  is the vector of length  $r$  of random effects and  $\varepsilon_k$  is the vector of length  $m$  of measurement errors.

$$Y_k = X_k\beta + Z_kb_k + \varepsilon_k \quad (4)$$



Observations from different studies are assumed independent, and within a study the random effects and measurement errors are also assumed independent. The measurement errors are assumed to follow a multivariate normal distribution of dimension  $m$ , namely  $\varepsilon_k \sim MVN_m(0, \mathbf{S}_k)$ , where  $\mathbf{S}_k$  is a  $m \times m$  covariance matrix. The random effects are assumed to follow a multivariate normal distribution of dimension  $r$  with  $b_k \sim MVN_r(0, \mathbf{A})$ , where  $\mathbf{A}$  is a  $r \times r$  covariance matrix. Note that the covariance matrix for the measurement errors  $\mathbf{S}_k$  is study specific (and so accounts for the within study correlation), and the covariance matrix for the random effects  $\mathbf{A}$  is common across the included studies (and so represents the variability between data from different studies). The structure assigned to  $\mathbf{A}$  defines how the random effects influence each other, options for this matrix include unstructured, autoregressive and compound symmetry amongst others. Ishak et al [32] note that ideally  $\mathbf{A}$  would be unstructured, however that as data is often limited in AD meta-analyses, this may lead to problems with estimating parameters correctly.

Ishak et al [32] discuss cases of equation (4) with a random intercept and a random slope. In these models they assumed that the study specific covariance matrix  $\mathbf{S}_k$  for the measurement errors  $\varepsilon_k$  was diagonal. Where  $\mathbf{S}_k$  was assumed diagonal, Ishak et al note that  $\mathbf{A}$  must not be diagonal (e.g. allowing  $\mathbf{A}$  to be unstructured, or following an autoregressive or compound symmetrical structure), as if both  $\mathbf{A}$  and  $\mathbf{S}_k$  were diagonal, measurements made at different time-points would be treated as independent. In the last model Ishak et al discuss, they do not restrict  $\mathbf{S}_k$  to be diagonal (although they recommend various approaches to simplify the estimation of  $\mathbf{S}_k$  such as using measures stated in the study report for the variances, keeping the correlations constant across studies or assuming a correlation structure that only requires one parameter to be estimated, such as an autoregressive or compound symmetry structure).

Ishak et al [32] note that the models that account for correlation between the observations in the studies provide better fits to the data than models that assume the observations are independent. Another benefit of this model is that as included studies are unlikely to be measured at exactly the same time points, modelling the correlation between observations allows these missing values to be treated as missing at random (see Rubin [42], and Little and Rubin [43] for definitions of missing data mechanisms).

However, these methods still have drawbacks. Ishak et al [32] noted that as they did not know the true values of the covariance parameters (as they used Aggregate Data only) they could not confirm how accurate the methods they proposed were. Jones et al [31]

presented methods for AD-MA as well as for IPD-MA, for cases where correlations within studies were reported or were missing. If sufficient information is available, they suggest using the second stage of two-stage IPD-MA of longitudinal data. If covariances between parameter estimates are not available, they recommend a sensitivity study to assess the robustness of conclusions to different correlation levels.

Similarly to Ishak et al [32], Maas et al [44] used multi-level modelling to perform meta-analysis of AD from longitudinal studies. They highlighted the benefits of modelling aggregate data using the model, such as inclusion of explanatory variables at different levels and exclusion of fewer studies from the analysis.

Peters and Mengersen [45] approach the issue of AD-MA of longitudinal studies from a slightly different direction. They argued that the aim of the meta-analysis, and the type of data available in the study reports, dictated the type of information required from included studies. The report defined five types of longitudinal AD-MA; (i) relevant time point MA where all available data at one time point is analysed (the time point being chosen for clinical relevance or to maximise available data), (ii) first/final time point MA (which uses only data from the first or final time point of included studies), (iii) all time points MA (where data at all time points are analysed, and data from different time points are considered independent, regardless of the fact that the same study can contribute to multiple meta-analyses), (iv) trend meta-analysis (where regression is used to model trend over time), (v) change in time point meta-analysis (where either the change from baseline to each time-point is analysed or the change between successive time-points is analysed).

The fourth method (trend meta-analysis) mentioned by Peters and Mengersen [45] requires either a slope estimates with variance to be supplied by the included studies, or the mean slope estimates with their variances at each time point. They suggest combining the study specific slope estimate using a random effects meta-analysis model, or if means and variances are reported at each time point within a study, to fit a regression model to each study and then combine the study specific slope estimates. The fifth method (change in time point meta-analysis) requires some summary estimate (for example the mean slope estimate) and a variance to be reported at each time point. The differences between the summary estimates at successive time points are then calculated and these differences are pooled between studies for each time point. Both the trend meta-analysis and change in time point meta-analysis account for the correlation inherent to longitudinal data by either by pooling slopes (which account for the trend over time) or by assessing changes over

time. However, the other three methods suggested by Peters and Mengersen [45] treat each time point as independent, ignoring the correlation between time points.

Ahn et al [46] focused on providing methods to model the correlations between data in longitudinal AD-MA, focusing on dose response models in the software NONMEM. The report states multiple levels of random effects (at either the study level or patient level) or correlated residual errors or both as possible methods to account for correlation. This is in line with the previous suggestions for longitudinal AD-MA [32].

#### *1.3.3.1 Limitations for longitudinal AD-MA*

For AD-MA, longitudinal studies often do not consistently report the summary measures needed for a particular method. For example, variances and covariances are not commonly reported [31], or may only be reported for specific time points. The lack of information can lead to issues with fitting models, for example Ishak et al [32] noted issues with convergence when estimating the covariance matrix, especially for structures that required multiple parameters to be estimated (for example unstructured covariance matrices). Methods for the imputation of missing variance data are described by Dakin et al [47] for Bayesian network AD-MA of longitudinal studies, and more generally by Boucher [48] for MA who focussed on modelling standard deviations over time for pain measurement. Jones et al [31] note that use of IPD solves the problem of non-reporting of covariance parameters, however this then requires a more costly IPD-MA rather than a less resource intensive AD-MA.

Additionally if longitudinal MA methods are used that only conduct a meta-analysis at specific time points, studies are excluded if they do not supply information at these specified time points – this can bias the conclusions drawn from the meta-analysis. Therefore, it is better to use AD methods that allow information across all possible time points to be used (as in equation (4), [32]), or to seek IPD from all studies concerned.

#### *1.3.4 Comparison of IPD-MA and AD-MA of longitudinal data*

Across the reports dealing with methods for longitudinal MA, there is a general consensus that the best approach is to obtain IPD, but that this is not always possible. Since ignoring correlation in longitudinal data can lead to unreliable parameter estimates and bias, IPD is preferred over AD to allow the correlation in the data to be properly modelled [31, 45], especially as relevant study reports often do not report the AD necessary to correctly model the correlation structure [32]. However, the time and resource expenditure problems of obtaining IPD datasets remain for longitudinal MA.

Overall, IPD methods for longitudinal MA can be considered a straightforward extension of the models generally used to model single study longitudinal data. A wide range of AD methods have been presented, however issues with them include a lack of reporting of variance and covariance parameters, ignoring correlation inherent to longitudinal data and differences between studies in measurement and follow-up times. Random effects, a common part of longitudinal data analysis, present a potential problem for meta-analysis due to the lack of data to estimate them, and the decisions that must be made as to what random effects to include. In general, for IPD versus AD meta-analyses, using IPD allows more flexible techniques to be used, as well as giving the potential for further investigation of specific subgroups and modelling structures (random effects and differing correlation structures).

### 1.3.5 Conclusions

A range of methods are presented in the literature both for AD-MA or IPD-MA. As IPD methods allow a wider choice of modelling methods as well as the opportunity to investigate the effect of individual level covariates on treatment effect, this thesis assumes the availability of IPD for studies included in the meta-analysis. In addition, the thesis will focus on the use of linear mixed effect models to represent the longitudinal data.

## 1.4 Time-to-event data

Time-to-event (or survival) data results from studies that aim to investigate the time between an individual entering the study, and experiencing a specific event (e.g. treatment withdrawal or death). Whilst some individuals may experience the event, others may reach the end of the study without experiencing the event, or may exit the study early for reasons unrelated to the event (e.g. moving away from the study location). Therefore, time-to-event data typically consists of two recorded values per study participant – a time (sometimes called a survival time, the minimum of the individual's event and censoring times) and a censoring indicator. The censoring indicator records a value of 1 if the individual experienced the event of interest, or a value of 0 if the individual was censored (dropped out for reasons unrelated to the event, or did not experience the event during the study period). An overview of time-to-event analysis is given in Collett [49].

There are three main types of censoring; left censoring, right censoring and interval censoring. Right censoring is defined as when it is known that the event occurs at some point after a certain measurement time (such as the event occurring after the end of the

study). Left censoring is defined as the event occurring before a certain time, but again it is unknown at what point before this recorded time the event occurs (such as the act of contracting a disease). Finally interval censoring is where the event is known to have occurred between two time points, for example between two follow-up times. The data considered in this thesis is right censored.

As time-to-event data is often skewed, and it has the unique property that both a time and a censoring variable is recorded, special methods are required to model the data.

Throughout a time-to-event analysis, interest lies in the risk or hazard of individuals experiencing the event at a given time. A hazard function (also known as the hazard rate, instantaneous death rate or intensity rate) is calculated, represented by  $\lambda_i(t)$ , which is interpreted as the risk of the individual in question experiencing the defined event in the next time interval, given that they have survived thus far (the formal definition is given in equation (5), where  $T_i$  is the survival time of the individual  $i$ ,  $t$  is a given time in the study, and  $\delta$  is a small number).

$$\lambda_i(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t \leq T_i < t + \delta t | T_i \geq t)}{\delta t} \right\} \quad (5)$$

Various models for the hazard function have been suggested to link the covariate values observed for an individual to their hazard of an event. A commonly used model in time-to-event analysis is the Cox Proportional Hazard (PH) model [50]. This is a semi-parametric model as no particular distribution is assumed for the survival times (a parametric model would be one that assumes the survival times are for example realisations from a Weibull distribution). This property of the model is often referred to as having an unspecified baseline hazard. The proportional hazards assumption of this model is that the hazard of experiencing the event in one group is proportional to the hazard of experiencing the event for a similar individual in a different group. This assumption can be checked by plotting the hazards of the two groups – if the hazard functions do not cross then the proportional hazards assumption is valid.

The equation for the hazard function for an individual  $i$  for a Cox PH model is given in equation (6).

$$\lambda_i(t) = \lambda_0(t) \exp(x_{2i}(t)' \beta_2) \quad (6)$$

In equation (6),  $\lambda_0(t)$  is the unspecified baseline hazard function (the function for the underlying risk of an event for all study participants), and the exponential function contains

covariates  $(\mathbf{x}_{2i}(t)')$  and their coefficients  $(\boldsymbol{\beta}_2)$  that adjust the baseline hazard function to the value of that individual's hazard at time  $t$ . The baseline hazard function is defined as the hazard function for an individual for whom all the values of the covariates held in the exponential function are zero [49].

To investigate the effect of different explanatory covariates on the hazard function for an individual, these explanatory variables can be included inside the exponential function of equation (6). If these explanatory variables are time stationary, there is no issue with inclusion of these variables in the equation. However, if these covariates are time variable, depending their type of variability over time, issues can arrive with their inclusion in the model.

There are two types of time dependent variables, namely exogenous (external) variables, and endogenous (internal) variables (see Kalbfleisch and Prentice [51] for formal definitions). Briefly, an exogenous variable may be considered predictable, for example the season of the year. Another type of exogenous variable may be described using a stochastic process independent to the individuals included in the study (such as air pollution count). In all cases, exogenous variables are independent of the time the event of interest occurs, and so can be included in the time-to-event model without issues.

However, endogenous variables rely on the existence of the study individuals to be measured. Examples include clinical measures and biomarkers. Various issues arise with endogenous variables with their inclusion in time-to-event models [52]. Endogenous variables are generally recorded with measurement error (for example when measuring blood pressure of an individual, measurements may vary due to biological and measurement factors, although exogenous variables can also be subject to measurement error). Secondly, as the endogenous variable relies on the existence of the individual the occurrence of the event implies that the endogenous variable can no longer be measured. Finally the endogenous variables are only measured at specific intervals across the study period – their values between the measurements cannot be stated with absolute certainty, unlike the exogenous variables whose values are predictable. Whilst exogenous variables can be included in the analysis of time-to-event outcomes through use of the extended Cox models (see Andersen and Gill [53], Fleming and Harrington [54], and Andersen et al [55]), including endogenous variables in a Cox model is not generally appropriate as doing so assumes that the values of the time dependent variable remain constant between

measurements (i.e. a step function). Therefore other methods are required, such as those described in Section 1.6.

Another type of term that may be included in the exponential function of equation (6) are frailty terms (the name given to random effects included in time-to-event analyses). These allow the population hazard model to be adjusted for each individual, and for the researcher to assess how much variation there is within the population for certain characteristics. Inclusion of frailty terms in time-to-event models is extended through the joint modelling framework, discussed in Section 1.6.

Two other important areas of time-to-event analyses are the extensions to allow for recurrent events and competing risks. Under certain conditions the event of interest in a time-to-event analysis may not be terminating; the individual may continue to survive after the event, such as time until treatment withdrawal. An extension of this is the case where the event is not terminating and can occur multiple times, for example asthma attacks. An individual could record multiple times until asthma attacks in one study, with time set to zero after each attack. Methods exist to account for this, however during this investigation it is assumed that events can only occur once in the study of interest, and additional data is not available for the study participant in question for any time point after they have experienced the event of interest.

The second extension (competing risks) involves single events rather than recurrent events, but allows for the fact that there may be multiple reasons that the event could occur. For example, when testing a new treatment if the outcome of interest is time until death, there are many causes of death that an individual could experience (such as cancer, cardiac arrest, etc.). It may be of interest to compare the hazards for the event of interest due to these separate reasons between the two treatment groups. This has led to competing risks models, where a separate hazard function is calculated for each event reason.

Apart from the Cox PH model, various other modelling approaches exist for time-to-event data, such as the Accelerated Failure Time (AFT) model. Unlike the Cox PH model (which assumes that the baseline hazard functions of the two treatment groups for example are proportional), the AFT model assumes that the hazard functions for both treatment groups (for example) are identical in shape but scaled differently over time. Different groups progress at different speeds along the graph of the hazard function, making the AFT model a suitable choice where progression through stages of a disease is of interest. One of the

joint modelling techniques described in Section 1.6 allows either Cox PH models or AFT models, although interest in this thesis centres on the Cox PH model.

In conclusion, the unique structure of time-to-event data requires specific methods to be used to analyse it. However, issues arise with time-to-event investigations when time dependent variables generated by internal processes of the study participant are potentially relevant explanatory variables. Joint modelling methods that allow the researcher to analyse time-to-event data including time dependent variables as explanatory variables are discussed in Section 1.6.

## 1.5 Meta-Analysis of Time-to-Event Data

The unique structure of time-to-event data has led to the development of methods specific to the data type. A range of references have extended these single study methods to a meta-analytic setting. This section details some of the issues encountered during the MA of time-to-event data, as well as a description of methodology for AD-MA and IPD-MA of time-to-event data.

### 1.5.1 Issues surrounding Time-to-Event Data Meta-Analysis

A range of issues surround the meta-analysis of time-to-event data. For example, for AD-MA, hazard ratios may not be reported at the same time points across different studies, meaning that only a subset of studies could contribute to a meta-analysis. Time-to-event data is complex with a unique structure, and so to ensure it is correctly modelled, it is often better to seek IPD. Again, however, this can lead to a range of issues including obtaining data, and the statistical expertise necessary to analyse it. However, a range of methods for both IPD and AD-MA of time-to-event data have been presented in the literature, some of which are now discussed.

### 1.5.2 Individual participant data meta-analysis of time-to-event data

#### 1.5.2.1 *One-stage methods for Time-to-Event Data MA*

Various modifications of the Cox PH model have been suggested to allow its use in a time-to-event meta-analysis. A key reference for the one-stage frequentist meta-analysis of time-to-event data is Tudur Smith et al [56] (also Tudur Smith [57]). Tudur Smith et al examine a range of 5 one-stage models applied to IPD time-to-event data. These models have the following formulations:



$$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{0k} + \beta_1 x_{1ki}) \quad (7)$$

$$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_1 x_{1ki}) \quad (8)$$

$$\begin{aligned} \lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{0k} + \beta_{1k} x_{1ki}) \\ \beta_{1k} &= \beta_1 + b_{1k} \\ b_{1k} &\sim N(0, \tau^2) \end{aligned} \quad (9)$$

$$\begin{aligned} \lambda_{ki}(t) &= \lambda_{0k}(t) \exp(\beta_{1k} x_{1ki}) \\ \beta_{1k} &= \beta_1 + b_{1k} \\ b_{1k} &\sim N(0, \tau^2) \end{aligned} \quad (10)$$

$$\begin{aligned} \lambda_{ki}(t) &= \lambda_0(t) \exp(b_{0k} + \beta_1 x_{1ki}) \\ \beta_{1k} &= \beta_1 + b_{1k} \\ b_{0k} &\sim N(0, \sigma^2) \\ b_{1k} &\sim N(0, \tau^2) \end{aligned} \quad (11)$$

In the above equations, the hazard function for the  $i$ th individual in the  $k$ th study is given by  $\lambda_{ki}(t)$ , with  $k$  ranging from 1 to  $K$  (the number of included studies), and  $i$  ranging from 1 to  $n_k$  (the number of individuals in the  $k$ th study). Treatment allocation is given by the  $x_{1ki}$  variable (as in the longitudinal MA section only a control and experimental treatment are considered, taking values 0 and 1 respectively).

The models presented provide a range of ways to account for between study heterogeneity in a time-to-event MA. For example, Equations (7) and (9) identify study membership using the fixed parameter  $\beta_{0k}$ , treating the first included study as the baseline ( $\beta_{01} = 0$ ). This term quantifies the difference in log hazard ratio between the baseline study and study  $k$ . The parameter  $\beta_1$  quantifies the difference in log hazard ratio for experiencing an event for those assigned to the experimental treatment compared to the control treatment.

Between study heterogeneity can be tested for by estimating study specific treatment parameters, and comparing the two models. The models shown in equation (8) and (10) demonstrate a baseline hazard stratified by study, rather than common across studies. This method accounts the log hazard ratio for risk of an event to completely differ in shape across time between studies. Models shown in equations (9) and (11) introduce random effects for the treatment effect, and the model in equation (11) includes a random effect  $b_{0k}$  that adjusts the baseline hazard between studies.

Various advantages and disadvantages are attached to the models shown in equations (7)-(11). The models with common baseline hazard across studies assume proportional hazards across all studies included in the meta-analysis, an assumption that may not be reasonable especially if the demographics or design of the included studies vary. The models with baseline hazards stratified by study assume proportional hazards only within studies,

potentially a more reasonable assumption. Models that rely on fixed effects to account for differences in the log hazard ratio for different groups between studies could become cumbersome as the number of studies increases, but they provide study specific parameter estimates. Inclusion of random effects allows the variation between studies attributable to treatment effect or from other sources to be quantified, by parameters  $\tau^2$  and  $\sigma^2$  respectively. However, random effects do not automatically produce study specific estimates of the parameters, although meta-analyses aim to provide overall rather than study specific estimates.

After comparisons of the models, Tudur Smith et al [56] noted that of these five models, the models that stratify baseline hazard by trial are more suited for meta-analyses as they preserve the multiple study structure. However, they note that with larger numbers of included trials, estimates from models with fixed trial effects or with stratification by trial could be unreliable. Random effect models are stated to be beneficial compared to a stratification by study approach if the meta-analysis consists of many included studies, each with a small sample size, but stratification is preferable to random effects modelling when there are few included studies with larger sample sizes.

Various other references also discussed IPD-MA of time-to-event data. Crowther et al [58] examine multilevel parametric time-to-event models for application to clustered time-to-event data (such as multi-centre or meta-analytic data), which use random effects to account for between cluster variation. An alternative to Cox PH models is presented by Crowther et al [59], who describe the one-stage IPD-MA of time-to-event data using Poisson regression models.

A summary of available methods for the IPD-MA of time-to-event data is provided by Katsahian et al [60]. Both Katsahian et al [60], and Michels et al [61] argue that inclusion of random effects is a valuable method to account for between study heterogeneity in time-to-event IPD-MA, an area also discussed by Rondeau et al [62].

#### *1.5.2.2 Two-stage methods for Time-to-Event Data MA*

For two-stage MA of time-to-event data, Thompson et al [63] used a two-stage method where Cox PH models were fitted separately to each included study. The model fitted to each included study was stratified by gender and randomised group, but included exposure and other covariates as model terms rather than additional stratification factors. They extracted log hazard ratio of the effect of a unit increase in exposure on the response, adjusted for the included covariates, and suggested combining these log(HR)s using a

random effects meta-analysis or a fixed effect meta-analysis. However, Thompson et al [63] report that one-stage and two-stage methods often present similar results.

Thompson et al [63] noted that whilst they utilised a two-stage model, a one-stage model similar to those discussed in Tudur Smith et al [56] (Cox PH models with random effects to account for between study heterogeneity) would be preferable. The main reason stated for not employing this approach was computational problems given the large size of the dataset (1.2 million participants over 116 included studies). However, they additionally note that use of two-stage models allows adjustment for different covariates between the included studies (which is beneficial if the covariates reported differ between included studies).

### 1.5.3 Aggregate data meta-analysis of time-to-event data

As with other data types, the use of AD-MA methods for time-to-event data can be valuable when IPD-MA is cannot be obtained from all studies. However, care must be taken to ensure that the methods used to analyse the time-to-event data in each study were appropriate (e.g. taking into account censoring).

Parmar et al [64] describe a range of methods available to extract hazard ratios (HR) from trials, (including methods for if the  $\log(\text{HR})$  and variance are reported, if  $\log(\text{HR})$  are reported along with a confidence interval, if the  $\log(\text{HR})$  are reported along with the  $p$  value of a log rank test under certain conditions) for use when estimating the  $\log(\text{HR})$  and variance from survival curves, or to estimate  $\log(\text{HR})$  and variance from survival curves whilst incorporating numbers at risk. These methods were then applied and investigated by Tudur Smith et al [65]. The methods for AD-MA of time-to-event data are also described by Tierney et al [66] in a manner accessible to non-statisticians.

Various other papers make key contributions to the AD-MA of time-to-event data. Williamson et al [67] describe methods to improve the estimation of  $\log(\text{HR})$  from survival curves or life tables, note that different approaches to extract data may be required across studies identified in the meta-analysis. Duchateau et al [68] state that the number of events should not be estimated from Kaplan-Meier curves unless the number of patients lost to follow up or censored is very small, and the number of patients at risk in the groups at the time at which the number of events is to be estimated is large. Arends et al [69] describe a method to perform AD-MA of survival proportions measured at multiple time points through a multivariate random effects model. In addition, Bennett et al [70] compare frequentist and Bayesian approaches to the AD-MA of time-to-event data.

However, the data available to conduct a time-to-event MA with aggregate data can often be limited, and drawing conclusions about treatment effects at the patient level on the basis of inferences based on aggregate data can be misleading [56].

#### 1.5.4 Comparison of IPD and AD MA of time-to-event data

Duchateau et al [71] compare IPD-MA and AD-MA and note that the results can differ, potentially due to the fact that IPD-MA will include all follow up in the analysis, whereas AD-MA will be based on information at only a subset of time points. In addition, an IPD-MA of time-to-event data allows a more effective investigation into the causes of heterogeneity than an AD-MA [56], which reduces the chances of conclusions drawn from the investigation being misleading.

#### 1.5.5 Conclusions

The methodology discussing the meta-analysis of time-to-event data is extensive. This thesis focuses on methods that rely on estimating hazard ratios from the data, although other methods also exist. For example, Barrett et al [72] suggests use of percentile ratios in place of hazard ratios in the second stage of a two-stage MA of time-to-event data, whilst Siannis et al [73] examine their use in a one-stage model. However, as these methods are not the focus of this thesis, they are not considered further.

Whilst many methods exist to meta-analyse time-to-event data, there is a trend towards the principle that availability of IPD time-to-event data is beneficial, as data at all time points rather than a subset can be analysed. A range of methods exist for IPD-MA of time-to-event data, including one-stage and two-stage models. Between study heterogeneity in one-stage models exist can be modelled in various including fixed or random terms, or stratification of the baseline hazard by study. The methods in this thesis will extend those for multi-study time-to-event data to multi-study joint data.

### 1.6 Joint models for longitudinal and time-to-event data

Joint models for longitudinal and time-to-event data (also referred to as joint models) simultaneously model longitudinal and time-to-event outcomes. Models of this type are typically used for one of three reasons; to account for informative dropout in a longitudinal study, to include a time dependent (longitudinally measured) explanatory variable in a time-to-event analysis, or to investigate possible association between a longitudinal and a time-to-event outcome where both are of equal interest [30]. Throughout this thesis, data

that contains both a longitudinal component and a time-to-event component of interest is referred to as joint longitudinal and time-to-event data, or more concisely joint data.

The use of joint models is increasing [74-76], with a growing number of methodology papers available in the literature describing joint model formulations for a range of scenarios. Joint models fall into two main groups, shared parameter models [74, 75] (which link models for the longitudinal and time-to-event outcomes by sharing parameters between them), and latent class models [76] (where the population being modelled are assumed to fall into a set of latent classes). This thesis concerns only shared parameter joint models, and any joint model discussed, unless otherwise stated, is a shared parameter joint model.

During this section an overview is given of various areas of single study joint modelling methodology including the general structure of the joint model, association structures within the joint model, software and model fitting, and finally the benefits and limitations of joint models.

#### 1.6.1 General Structure of a Joint Model

A simple joint model contains a longitudinal sub-model and a time-to-event sub-model, which are linked through an association structure. The association structure quantifies the dependence between the longitudinal and the time-to-event outcomes.

A wide range of joint models have been presented in the literature. Proposed longitudinal sub-models include mixed effects models, use of splines, and non-linear models. Suggested time-to-event sub-models include proportional hazards (PH) models, AFT models, or parametric models. An overview of joint modelling methodologies is given by Gould et al [75] and Davidian et al [74], with useful textbooks by Rizopoulos [52] and Elashoff et al [77]. Other key references of joint models include Wulfsohn and Tsiatis [78] (generally regarded as the first joint modelling paper) and Henderson et al [30].

During the thesis, the joint models take following structure:

$$\begin{aligned} Y_i &= X_{1i}\beta_1 + Z_i b_i + \varepsilon_i \\ \lambda_i(t) &= \lambda_0(t) \exp(X_{2i}\beta_2 + W_{2i}(t)) \\ W_{2i}(t) &= g(X_{1i}, \beta_1, Z_i, b_i) \end{aligned} \tag{12}$$

In equation (12), the longitudinal outcome is modelled using a mixed effects model (first line), and the time-to-event outcome through a PH model with an unspecified baseline hazard (second line). The longitudinal and time-to-event sub-models are linked using term

$W_{2i}(t)$ , which takes some function  $g$  of the longitudinal sub-model and inserts it into the time-to-event sub-model (third line in equation (12)). Exact model formulae used in this thesis are stated as required.

### 1.6.2 Association Structures

A range of association structures (also called sharing structures) for joint models have been proposed in the literature, each of which allows different patterns of dependence. A good overview is given by Gould et al [75].

This thesis investigates models that employ a proportional association structure. In this structure, the random effects from the longitudinal sub-model are correlated with the frailty term in the time-to-event sub-model. As such, the  $W_{2i}(t)$  has the following format:

$$\begin{aligned} W_{2i}(t) &\propto W_{1i}(t) \\ &= \alpha W_{1i}(t) \\ &= \alpha \mathbf{Z}_i \mathbf{b}_i \end{aligned}$$

In the above equations, it is stated that the  $W_{2i}(t)$  is proportional to  $W_{1i}(t)$  (which consists of the random effects of the longitudinal sub-model). The zero mean random effects are represented by  $\mathbf{b}_i$ , with design matrix  $\mathbf{Z}_i$ , and association parameter(s)  $\alpha$ . In a single study case, there is a single  $\alpha$  parameter, however during the extension to a multi-study case, which permit random effects at multiple levels, separate association or  $\alpha$  parameters are permitted for each level of random effects (see Section 3.3).

The proportional random effects only association structure can also include separate random effects only structure, where separate association parameters  $\alpha_1, \alpha_2, \dots$  are permitted for each shared random effect. Other association structures involving both the fixed and the random effects are discussed in Rizopoulos [52]. The current value structure inserts the longitudinal sub-model (apart from the error term) with coefficient into the time-to-event sub-model, giving  $W_{2i}(t) = \alpha m_i(t) = \alpha(\mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}_i\mathbf{b}_i)$ . The current slope structure inserts the first derivative of the longitudinal trajectory (apart from the error term) with respect to time into the time-to-event sub-model with coefficient, giving  $W_{2i}(t) = \alpha m_i'(t) = \alpha \left( \frac{d}{dt} (\mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}_i\mathbf{b}_i) \right)$ . Association structures modelling the cumulative effect of the longitudinal trajectory integrate under the longitudinal trajectory utilise  $W_{2i}(t) = \alpha \int m_i(t) = \alpha \int (\mathbf{X}_i\boldsymbol{\beta}_1 + \mathbf{Z}_i\mathbf{b}_i)$ . A weighting function can also be included in the cumulative effect, to allow recent longitudinal measurements to have a greater effect on the risk of an event than older measurements.

These association structures each give association parameters with different interpretations [52, 75]. For the proportional random effects only association structure, the association parameter  $\alpha$  quantifies the effect of individual deviation from the population mean value at a given time point on their risk of an event. The separate random effects only structure has association parameters that quantify the effect of individual deviation from the population effect of the covariate assigned the random effect on the risk of an event. The current value structure quantifies the risk of an event for an individual based on their true current value of their longitudinal outcome. The current slope structure gives the risk of an event for an individual based on the current rate of change of their longitudinal outcome. The cumulative structure details how the history of an individual's longitudinal outcome up to a given time point affects their risk of an event.

The difference in interpretation of association parameters could affect the meta-analysis of results from joint models, an issue discussed in Chapter 2 and Chapter 3.

### 1.6.3 Currently available software

A range of software is currently available to fit single study joint models. In the statistical software R [79], there are two main frequentist packages to fit single study joint data, namely *joiner* [80] and *JM* [81]. These software packages differ in the models that they provide, this is further discussed in Section 5.1.1. The package *JMBayes* [82] provides a Bayesian approach to joint modelling in R. Whilst this thesis focusses on the R software, joint modelling is also possible in other software. The *stjm* function [83] in STATA [84] is well established, and in the SAS software [85] various options exist including the *PROC NLMIXED* function, as well as other specifically designed macros (see Chapter 2 for a review of current use of joint modelling software). However, there did not appear to be readily available software designed to handle multi-study joint data.

### 1.6.4 Discussion

One major benefit of joint modelling is that for a longitudinal study, it allows the investigator to model informative dropout, which if ignored could lead to biased results [86]. Additionally for a study where a time-to-event outcome is of interest, joint modelling allows the investigator to include a longitudinal covariate (that may not be continuously measured or may be recorded with error) in the time-to-event model. Additionally, Ibrahim et al [87] also notes that joint modelling of available longitudinal and time-to-event data leads to more efficient estimates of the treatment effects in both the time-to-event and the longitudinal sub-models, and gives a larger power for a smaller sample size in a study.

However, there are limitations to joint models due to their complexity. Expectation-Maximisation (EM) algorithms are commonly used during model fitting, but can be slow to converge, and does not automatically produce estimates of the standard errors [88]. Due to use of the unspecified baseline hazard, it is necessary to use bootstrap re-samples to obtain standard errors, as relying on the profile likelihood could lead to underestimation [89]. Whilst joint models reduce bias in certain cases, and allow inclusion of endogenous variables in time-to-event models, their complexity may cause some investigators to opt for more naïve but more straightforward methods such as the extended Cox model [53-55].

## 1.7 Meta-Analysis of Joint Data

The benefits of both subject areas of meta-analysis and joint models for longitudinal and time-to-event data are clear. Meta-analysis allows currently existing data to be pooled, allowing smaller effects to be identified, potentially reducing the number of trials required in the future, and thus reducing the number of patients exposed to experimental, potentially detrimental treatments. Joint models allow more efficient and in-depth investigation of data involving both longitudinal and time-to-event variables, whilst accounting for issues such as informative dropout in longitudinal data, or time varying covariates in time-to-event models. The meta-analysis of joint models would therefore provide a useful tool to both the statistical and medical research communities.

A review of the available methodology for separate meta-analysis of time-to-event data or longitudinal data identifies a range of considerations. For time-to-event data, it is preferable to have IPD for each study to allow the full follow-up of the data to contribute to the analysis. For longitudinal data, the multilevel nested structure of the data must be accounted for. Again IPD is preferred, but given sufficient reporting in available studies, AD-MA are feasible. However, given the issues that joint data MA may inherit from longitudinal MA, time-to-event MA or single study joint models, IPD is likely to be preferable.

A review of current literature on the meta-analysis of joint longitudinal and time-to-event data did not reveal literature or software specific to the area of meta-analysis. Brombin et al [90] provided an analysis of multi-centre joint data, however they did not present methods designed to deal with the meta-analysis of joint data, such as a comparison of methods to deal with between study heterogeneity. The methods are comparable, as both involve three level joint data, however meta-analytic and multi-study analyses differ in the number of units at each level. Meta-analyses involve few studies at the top level, each



including many individuals. Multi-centre studies involve many centres at the top level, typically each containing few individuals. As such, suitability of methods to account for differences between top level units may differ between multi-centre and meta-analytic analyses. In addition, generalisable software designed for joint data meta-analysis did not appear available in the literature. As such, there is a gap in methodology linking the important areas of meta-analysis and joint models for longitudinal and time-to-event data.

## 1.8 Thesis Aims and Structure

### 1.8.1 Motivation for research

In summary, this chapter introduced the basics of meta-analysis, and the two types of data of interest in this thesis (longitudinal and time-to-event). The joint model formulation for a single study case including various model structures and association types were outlined. An overview of the literature underpinning meta-analysis of longitudinal data and time-to-event data was given. The single reference that implemented a three level joint model was identified. The large gap in the methodology linking the important areas of longitudinal data, time-to-event data and meta-analysis was noted, which provides the motivation for this thesis. Methodology and software to ensure that longitudinal and event-time outcomes from multiple studies can be appropriately modelled is an under-developed area of research, especially given the increasing popularity of joint modelling.

### 1.8.2 Aims of Thesis

This thesis aims to extend current meta-analytic and joint modelling research in the following ways:

1. Develop methods for the two-stage IPD-MA of joint longitudinal and time-to-event data comparing separate and joint models through real data application, and a simulation study investigating the behaviour of methods under a range of different event rates, association parameters and levels of between study heterogeneity.
2. Develop methods for the one-stage IPD-MA of joint longitudinal and time-to-event data investigating methods to account for between study heterogeneity as well as comparing use of joint and separate models, through a real data application and a simulation study investigation varying association levels, number of included studies and levels of between study heterogeneity

3. Develop software in the R programming language [79] to aid researchers conducting MA of joint longitudinal and time-to-event data.
4. Assess current reporting standards of joint modelling investigations of single study medical or biostatistical data with a view to future AD-MA of published analyses.

### 1.8.3 Thesis structure.

In this first chapter of the thesis, I have discussed the background of the methodologies developed in this thesis and outlined the thesis aims. I now conclude the chapter with a discussion of the structure of the thesis.

Chapter 2 contains the results of a review of current reporting standards in the literature of joint modelling analyses of medical or biostatistical data, with an aim to assess the current feasibility of extracting sufficient information to conduct an AD-MA of the published results. A range of aspects of the identified studies including disease area, joint model type, and software employed were recorded, as well as an assessment of how many of the identified studies provided sufficient information to contribute to future AD-MA.

I develop and present methodology for the meta-analysis of joint longitudinal and time-to-event data in Chapter 3. Methods are separated into two-stage and one-stage methods. The two-stage methods present guidelines designed to describe the process that should be undertaken in a two-stage MA. The one-stage methods describe various groups of models designed to account for between study heterogeneity in several ways, including fixed parameters for study membership and interaction between study membership and other covariates of interest, study level random effects, and stratification of the baseline hazard by study.

The R software package that I have developed during this thesis to implement the methodology discussed in Chapter 3 is presented in Chapter 4. This package includes methods for the preparation and visualisation of multi-study joint data, functions to aid the second stage of a two-stage MA, and functions to fit one-stage meta-analytic joint models. The software is designed to be flexible for use in future multi-study investigations, and is freely available. The theory used by the functions in the package (e.g. to fit models) is discussed in Chapter 5.

A real data application of the one-stage and two-stage methods is presented in Chapter 6. This chapter involves the analysis of the INDANA [91] dataset, a multi-study dataset of hypertensive patients which contains the longitudinal outcome Systolic Blood Pressure

(SBP), and time-to-event outcomes time to death, time to myocardial infarction and time to stroke. I compare meta-analyses based on joint methods to meta-analyses based solely on longitudinal or on time-to-event methods, in order to assess for this real dataset whether a benefit to joint models exists. This analysis is designed to test the methodology developed in a real data setting, rather than influence clinical practice.

A simulation study is presented in Chapter 7. Separate simulation studies are undertaken for two-stage methods and one-stage methods. In both cases, meta-analyses using joint models are compared to meta-analyses using longitudinal or time-to-event models to assess any benefit in using the more complex joint modelling methodology. During the two-stage MA simulation study, a range of scenarios including varying association level, event rate and between study heterogeneity in treatment effect are investigated. During the one-stage MA simulation study, a range of associations are tested for a single event rate. The effect of varying between study heterogeneity, and varying numbers of available studies on the estimation of one-stage model parameters are also assessed.

A discussion of the work presented as well as an assessment of its impact on future research is given in Chapter 8. This chapter also includes a concise statements of conclusions drawn from the research, as well as a presentation of planned future work.

## Chapter 2 : Review of reporting of joint models in the literature

---

Joint modelling of longitudinal and time-to-event data is an area of increasing research [75, 76, 92]. As such, results of joint longitudinal and time-to-event analyses could be expected to be increasingly published in the literature. Aggregate Data Meta-Analyses (AD-MA) numerically pool study level results to obtain an overall estimate of, for example, treatment effects. These study level results can be extracted from publications, or obtained through direct contact with publication authors. Whilst not the gold standard for meta-analyses, AD-MA are commonly performed in place of Individual Participant Data Meta-Analyses (IPD-MA), as they can be less resource intensive to complete. However, they are reliant on sufficient information being available from publications or authors to conduct the meta-analyses. This chapter describes a review conducted of the published results of joint longitudinal and time-to-event analyses of medical and biostatistical data, which aims to assess whether AD-MA of joint models is feasible given current reporting practices. This review has been published in Sudell et al [93].

The chapter starts with an examination of the issues to be considered when assessing the current feasibility of AD-MA of joint model results. The structure of the conducted review is then stated, followed by the results. The chapter concludes with a discussion of AD-MA of joint data and recommendations for future reporting of joint longitudinal and time-to-event models in the literature.

### 2.1 Motivation for the review

An AD-MA using published data would only be successful if the studies identified during the meta-analysis search state results in a way that is useable in a meta-analysis. It is assumed that for an AD-MA of joint longitudinal and time-to-event models, separate meta-analyses using standard methods would be performed for the parameters from the longitudinal sub-model, from the time-to-event sub-model and from the association structure.

The association parameters are assigned a separate class to the time-to-event sub-model fixed effect parameters as their interpretation differs. The association parameters quantify the effect of some component or function of the longitudinal sub-model on the risk of an event. They communicate both the effect of the function of the longitudinal outcome on

risk of an event, but also the relationship between the longitudinal and time-to-event outcomes, an important aspect as if the relationship is non-significant, the difference between results from separate longitudinal or time-to-event models, and from joint models, may be small. The time-to-event sub-model fixed effect parameters quantify the direct effect of variables that do not change over time on the risk of an event. As such, separating the MA for the time-to-event fixed effect coefficients and the association parameters highlights their difference in interpretation. Additionally, performing MA for the association parameters communicates whether a significant relationship between the longitudinal and time-to-event outcomes exists.

To perform the meta-analyses, the sample size of the study must be reported, as well as the parameter estimates with some representation of their precision (standard error, or confidence intervals with stated significance level). If interest also lies in results from particular sub-groups, this same information must be available for each sub-group of interest.

A wide range of types and structures of joint models are currently available to researchers. Section 3.2.4, discusses that pooling results from different types of joint models may not have meaningful interpretations. For example, a joint model that uses a random effects only proportional association structure has an association parameter that is interpreted as the effect of individual deviation from the population mean trajectory on the risk of an event [75]. However, a joint model that uses a current value association structure has an association parameter that quantifies the effect of the current recorded value of the longitudinal outcome on the risk of an event [52]. Pooling the association parameters from such different joint modelling structures would not be interpretable, and so care needs to be taken during AD-MA not to make this mistake.

An additional consideration stems from the choice of association structure. If a random effects association structure is used, then fixed effect parameter estimates can be directly read from each sub-model. However, if an association structure that involves both fixed and random effects is used, then direct, indirect and overall effects for covariates can be extracted from the time-to-event sub-model (see Ibrahim et al [87] and Section 5.1.1). Consider a joint model that includes treatment assignment as a fixed covariate in both the time-to-event and the longitudinal sub-models, and that utilises a fixed and random effects current value association structure. The time-to-event sub-model will contain a direct effect of treatment assignment by inclusion of the variable as a fixed effect in the time-to-

event sub-model. It will also include an indirect effect, resulting from the inclusion of the treatment effect covariate as a fixed effect in the longitudinal sub-model, present in the time-to-event sub-model as a result of the association structure linking the sub-models. The indirect effect will be a product of the association parameter and the longitudinal treatment effect coefficient. An overall effect of the treatment assignment in the time-to-event sub-model is then the sum of the direct and indirect effects. In joint models of this or similar format, this overall effect is generally the quantity of interest as it fully describes the effect of the covariate on the risk of an event. As overall effects from joint models utilising association structures involving both fixed and random effects consist of both a direct and indirect effect, whereas the effects extracted from joint models utilising random effects only association structures are direct effects only, it may not be interpretable to combine parameter estimates from joint model analyses that used different joint modelling structures.

Consequently, as it may not be meaningful to pool estimates from joint models with different structures, it is important for the joint modelling structure (parameters included in the sub-models, and association structure) to be discernible from the study report.

An AD-MA may be difficult or impossible to perform if information such as the joint model structure is not provided for the contributing study in question, without either seeking additional information or full IPD datasets from the study authors. In order to assess the current standard of reporting of joint models, a review was conducted of papers published in the literature using joint modelling methods.

## 2.2 Review Methods

### 2.2.1 Identification of Papers

Medline, Pubmed and Scopus datasets were searched for studies using joint models for longitudinal and time-to-event data to analyse medical data (search strategies available in Appendix 1).

Any papers that mentioned joint models for longitudinal (or repeated measures over time) data and time-to-event (or survival, event time or event history) data were identified. The citations were downloaded and duplicates identified and removed. The abstracts and keywords of the remaining papers were then examined and any irrelevant papers were removed. Examples of disregarded papers include papers modelling body joints, those discussing joint models as a future extension or alternative to methods used, or papers

using two-stage approaches rather than simultaneous estimation of the longitudinal and time-to-event sub-models. Papers not relating to a medical or biostatistical dataset were also disregarded (for example application of joint models to data from plant or animal subjects except from those modelling human diseases input into animal hosts). In addition, papers involving repeated measures over space rather than time (e.g. repeated measures across different tumour sites recorded at the same time point) were removed. If the relevancy of identified papers was unclear from the abstract, the full text was obtained and examined after which the study was included or discarded.

Retained papers were sorted into an applied and methodological group. Whilst many papers from the methodology group presented application of methods to example datasets, these were considered re-analyses of data or demonstration of methodology rather than primary analyses aimed to influence future medical practice. Additionally, methodological papers could be expected to better report results than applied papers, as their authors tend to be joint modelling experts. This review aimed to assess the reporting of joint modelling analyses in the general medical literature, so the review focussed only on the applied group.

### 2.2.2 Data extraction

Throughout this chapter, references identified as relevant to the review are referred to as studies. Other publications (such as those cited by the identified studies) are termed papers.

The year of publication, author, journal, joint model type, association structure between the longitudinal and time-to-event sub-models, types of sub-models, Bayesian or frequentist methods, and software used were recorded from the identified studies. Additionally disease area was recorded with respect to the type of longitudinal and time-to-event data, for example studies modelling biomarkers in heart disease patients after a transplant operation were classed as transplant data.

Also the source of the methods used in the study was recorded. If the study developed methods specific to their dataset or research objectives, a value of “own methods developed” was recorded. If the study referenced a particular paper or papers as the source of the methods used in their investigation, the papers referenced were recorded.

The availability of each piece of information necessary for an AD-MA to be performed was recorded. Specifically, the number of participants, significance level, and the presence of

the parameters from the longitudinal and time-to-event sub-models, and the association structure along with estimates of their precisions were searched for in each study. The significance level was identified either through its direct statement in the study text, or in a statement of a confidence interval of a particular size on tables, graphs or footnotes specifically relating to the joint models fitted in the analysis. Feasibility of meta-analyses was assessed separately for the longitudinal, the time-to-event and the association parameters. For each group of MAs to be considered possible, the number of participants contributing to the analysis, the relevant parameter estimates and an estimate of their precision (their standard errors or a confidence intervals with accompanying significance level) had to be reported.

As noted earlier, it is assumed that an AD-MA of results from joint longitudinal and time-to-event models will involve separate MA for each parameter from each of three groups – namely the parameters from the longitudinal sub-model, the parameters from the time-to-event sub-model and the association parameters. Ideally MA should be possible for each of these three groups. However, if insufficient information is reported in a study, MA may only be possible for a subset of these three groups.

Joint models in single study cases may be utilised for a variety of reasons, for example to account for study dropout in a longitudinal study, to include a longitudinal covariate measured with error in a time-to-event analysis, or simultaneously analysing a longitudinal and a time-to-event outcome both of which are of interest [30]. If joint models were utilised in order to account for informative dropout in a longitudinal study, the study might not report the time-to-event sub-model parameters as the sub-model exists to account for a structural issue in the data rather than to model an outcome of interest. If a joint model was employed in order to include a longitudinal covariate measure with error as a time variable covariate in a time-to-event model, again the parameters of the longitudinal sub-model may not be clearly reported. Therefore the reason reported in the study for using joint modelling methods was recorded, to assess whether it affected the proportion of possible MA for each of the three groups (longitudinal, time-to-event, and association parameters).

The aim of this review was not to perform MA of the study level results of joint modelling analyses reported in the literature, it was solely to assess if such an MA were undertaken, what proportion of identified studies could currently contribute. It should be emphasised



that methods for AD-MA of study level results are comparable to the second stage of two-stage methods for IPD data, which are discussed in Chapter 3.

## 2.3 Results

Searches were conducted on the 15<sup>th</sup> September 2015. A flowchart depicting the process the review followed is shown in Figure 1. Once duplicate references and an erratum paper that corrected an author's name were removed there were 618 remaining references. Of these, 210 were classed as methodological papers, and 343 were disregarded for reasons mentioned above. In total, 65 studies remained [90, 94-157] that applied joint models for longitudinal and time-to-event data to medical or biostatistical datasets with the aim of influencing healthcare rather than demonstrating new modelling methods.

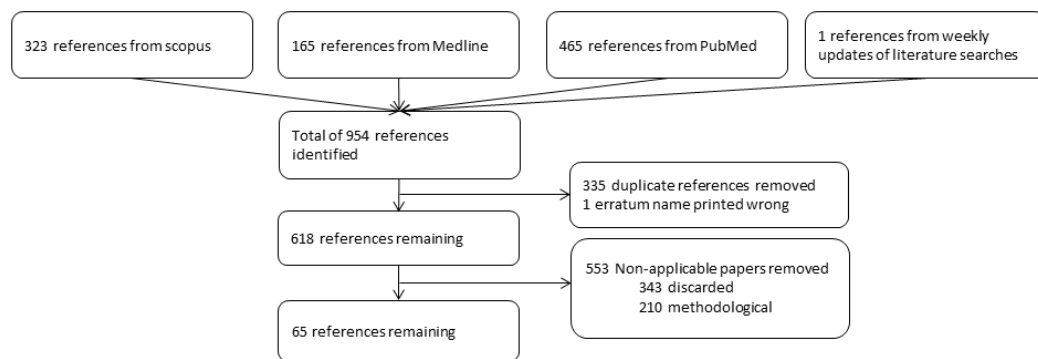


Figure 1: Flowchart of review of reporting of joint modelling of longitudinal and time-to-event medical or biostatistical data in the literature

### 2.3.1 Characteristics of identified studies

#### 2.3.1.1 Year of Publication

A plot of the years of publication of the identified studies is shown in Figure 2. The distribution of year of publication of the identified studies appears skewed towards more recent dates. The median year of publication was 2014, with interquartile range 2011 to 2014, range 2001 to 2015. An overall trend is noticeable of increasing numbers of publications of applied joint models per year, although there is some variation between the years and the maximum number of publications published in a year is only 20.

Numbered lines (1-6) are included on Figure 2 to indicate years when events or publications that could have influenced the rate of publication of applied joint modelling papers occurred. Line 1 indicates the publishing of the seminal paper by Wulfsohn and Tsiatis [78] commonly cited as one of the first joint modelling papers. Their methodology was extended by Henderson et al [30] in 2000 (line 2). Line 3 represents two papers of

interest, firstly by Tsiatis and Davidian [92] who published a review of joint modelling methodology, and secondly by Guo and Carlin [158], who published a paper giving examples of implementation of joint models in current software. In 2010 (line 4), Rizopoulos published a paper detailing the R joint modelling package *JM* [81]. Line 5 highlights 2012, which saw the publication of a joint modelling textbook [52], and papers describing joint modelling options in Stata (Crowther et al [159]) and the *joiner* package in R (Philipson et al [80]). Further papers were published by Crowther et al concerning joint modelling in Stata in 2013 (line 6, [83, 160]). In addition to these events, the number of joint modelling workshops, talks and conferences has increased in recent years (see <https://www.liverpool.ac.uk/translational-medicine/departmentsandgroups/joiner-workshops/>, <http://eur.academia.edu/DimitrisRizopoulos/Talks>, <http://www2.le.ac.uk/departments/health-sciences/research/biostats/staff-pages/mjc76>, accessed 10<sup>th</sup> February 2016). Although it is unclear by how these events or publications influenced the use of joint modelling methods, there is an apparent increase in the use of the methods after 2012.

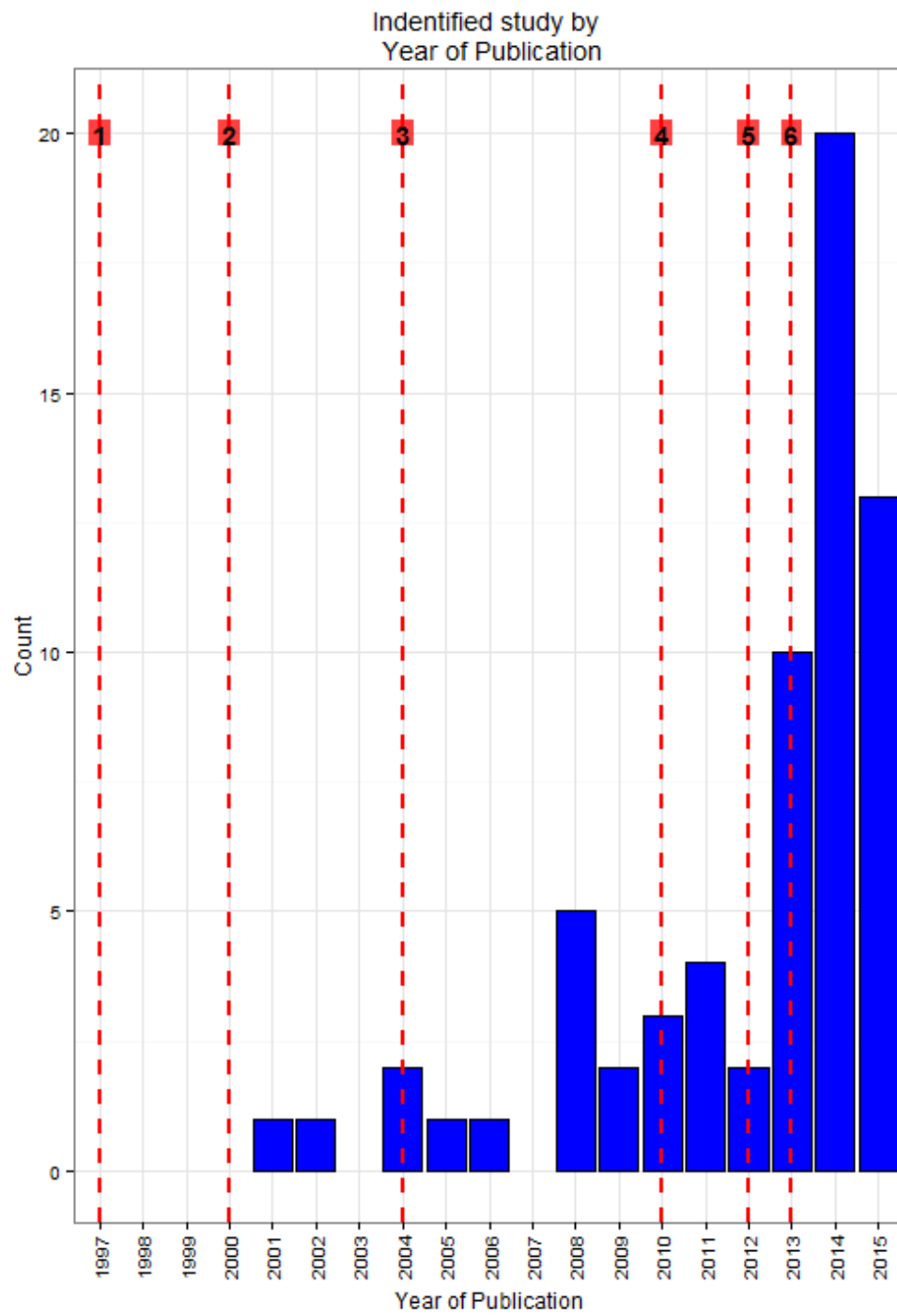


Figure 2: Year of publication of identified studies. Line numbers identify possibly influential publications (see main text)

Full text or abstract available		N (%)
Full text		63 (96.9)
Abstract		2(3.1)
Disease Area		
Cancer related data		10 (15.4)
HIV/AIDS		9 (13.8)
Patient status after transplants		8 (12.3)
Cognitive decline		7 (10.8)
Glaucoma		4 (6.2)
Renal disease		4 (6.2)
Disability in the elderly		3 (4.6)
Heart related data		3 (4.6)
Schizophrenia		3 (4.6)
Sclerosis		3 (4.6)
Other		11 (16.9)
Journal		
Statistics in Medicine		5 (7.7)
Journal of the Royal Statistical Society. Series C: Applied Statistics		4 (6.2)
Ophthalmology		3 (4.6)
Quality of Life Research		3 (4.6)
Journal of the American Geriatrics Society		2 (3.1)
Journal of the American Statistical Association		2 (3.1)
Journals of Gerontology - Series B Psychological Sciences and Social Sciences		2 (3.1)
Statistical Methods in Medical Research		2 (3.1)
Other (only one study per journal)		45 (64.6)
Reason for joint modelling use*		
To investigate the link between longitudinal and time-to-event outcomes		43 (66.2)
To account for dropout		22 (33.8)
To include longitudinally measured variable in time-to-event model		4 (6.2)
To increase efficiency		3 (4.6)
To reduce bias		2 (3.1)
Easier to interpret		1 (1.5)
To use all of available data		1 (1.5)

Table 1: Characteristics of Identified Studies (\*Note for "disease area" and "journal" only one value was recorded per included study giving total N=65, however for "reason for joint modelling use" multiple reasons could be recorded per included study giving total N≥65)

### 2.3.1.2 Full text or Abstract

Full texts were obtainable for 63 of the identified studies (96.9%) [90, 94-105, 107-138, 140-157], with only abstracts available for 2 studies (3.1%) [106, 139], see Table 1. Some individuals were listed as authors on multiple studies, suggesting that the group of individuals applying joint models may currently be limited.

### 2.3.1.3 Disease Area

A wide range of disease areas were reported in the identified studies (see Table 1). The most commonly reported types were Cancer (10, 15.4%), HIV/AIDS (9, 13.8%), transplant

data (8, 12.3%) and cognitive decline (7, 10.8%). The wide range of disease areas identified by this review demonstrates the applicability of joint modelling methods to a variety of areas of medicine. However, this also highlights the issue that currently identifying multiple studies utilising joint modelling methods in the same disease area for use in an AD-MA could be difficult.

#### *2.3.1.4 Journal*

The studies identified in this review were published in a range of journals, with 8 journals providing more than one identified study (see Table 1). This could indicate that currently there may not be a preferred journal to present joint modelling studies for medical or biostatistical datasets in.

#### *2.3.1.5 Reason for use of joint model*

Some studies identified in the review provided multiple reasons for using joint modelling methods. All reasons given by a study were recorded (see Table 1), giving a total number of reported reasons greater than the number of identified studies. The two most commonly reported reasons for joint model use were to investigate a link between the longitudinal and time-to-event outcomes, or to account for dropout in the study. Of the identified studies, only 4 stated that they used joint models in order to include a time varying covariate measured with error in a time-to-event analysis.

Source of joint modelling methods used	N (%)
Own methods developed	18 (27.7)
Guo-Carlin 2004 [158]	13 (20.0)
Rizopoulos 2010 ( <i>JM</i> R package) [81]	10 (15.4)
Henderson et al 2000 [30]	7 (10.8)
Tsiatis and Davidian 2004 [92]	7 (10.8)
Rizopoulos 2012 [52]	6 (9.2)
Wulfsohn and Tsiatis 1997 [78]	6 (9.2)
Diggle et al 2008 [161]	3 (4.6)
Crowther et al 2013 [160]	2 (3.1)
Proust-Lima et al 2009 [162]	2 (3.1)
Rizopoulos 2011 [163]	2 (3.1)
<b>Approach</b>	
Frequentist	45 (69.2)
Bayesian	17 (26.2)
Both	1 (1.5)
Unclear	2 (3.1)
<b>Association structure</b>	
<b>Fixed and Random Effects</b>	<b>33 (50.8)</b>
Current Value of Fixed and Random Effects	24 (36.9)
Current Slope (first derivative) of Fixed and Random Effects	3 (4.6)
Current Value of Fixed and Random Effects and Current Slope (first derivative) of Fixed and Random Effects	5 (7.7)
Fixed and random effects without covariates	1 (1.5)
<b>Random Effects only</b>	<b>27 (41.5)</b>
Intercept only	5 (7.7)
Random Effects with covariates	7 (10.8)
Random Effects without covariates	9 (13.8)
Random Effects unclear with or without covariates	6 (9.2)
<b>Latent Class</b>	<b>3 (4.6)</b>
<b>Specialist association structure</b>	<b>4 (6.2)</b>
<b>Unclear</b>	<b>4 (6.2)</b>

Table 2: Methods used in identified studies (Note for "Approach" only one value was recorded per included study giving total N=65, however for "Source of joint modelling methods used" and "Association structure" multiple values could be recorded per included study giving total N≥65)

#### 2.3.1.6 Source of methods used

In total, 18 (27.7%) of the identified studies used modelling methods specific to their study. The remaining identified studies used methods described in other papers, of which a total of 38 unique papers were cited. Ten of these papers were referenced by more than one of the identified studies (see Table 2). Some of these commonly cited papers were software specific. For example the papers of Rizopoulos 2010 [81] and 2012 [52] are R related, whilst Crowther et al 2013 [160] is Stata related. Other commonly cited papers provided

overviews of the methodology and implementation of joint models, such as Proust-Lima et al [162] and Guo-Carlin [158].

#### *2.3.1.7 Modelling approach*

For each included study it was recorded whether a Bayesian or Frequentist approach was undertaken (see Table 2). Of the 65 identified studies, a total of 45 (69.2%) undertook a frequentist approach, 17 (26.2%) a Bayesian approach, 1 (1.5%) used both approaches (in separate model fits) and in 2 (3.1%) studies the approach was unclear (these were the two abstracts). The larger proportion of studies taking frequentist approaches could be attributable to the fact that a larger number of papers and packages are based on frequentist methods. In addition, the original joint modelling textbook [52] deals with frequentist methods.

The type of model used for each of the longitudinal and the time-to-event sub-models was recorded. Overall, there were 21 unique model types recorded for the longitudinal sub-model, with unclear model type recorded for 1 study. The most common type of longitudinal sub-model specified was the linear mixed effects model (35 studies, 53.8%), followed by mixed effect models that used splines (6 studies, 9.2%), or mixed models with an unspecified structure (5 studies, 7.7%). Other modelling methods recorded included use of different mixed models dependent on latent class membership, non-linear models with or without splines, and models using change points.

A wide range of models were reported for the time-to-event sub-models, with 4 studies recorded as unclear. The most commonly reported time-to-event sub-model type was the Cox PH model (8 studies, 12.3%). However, other methods were also reported across the studies, including models with a parametric baseline (e.g. a Weibull PH model which was reported by 5 studies, 7.7%), PH models with piecewise constant baseline hazard (4 studies, 6.3%) or spline based baseline hazard (2 studies, 3.1%). Additionally there was mention of fully parametric models, including the Weibull (5 studies, 7.7%), and the exponential model (1 study, 1.5%). Additionally 1 study (1.5%) examined both Weibull and exponential models for the time-to-event sub-model.

#### *2.3.1.8 Association structure between longitudinal and time-to-event sub-models*

The association structure links the longitudinal and the time-to-event sub-models in a joint model. A variety of options are currently available to researchers, each with slightly different interpretations. The association structures reported in the studies identified by the review are reported in Table 2. Some identified studies fitted more than one joint

model in their analyses, with varying association structures, allowing a total of more than 65 (the number of identified studies) recorded association structures.

An association structure that involves some function of both the fixed and random effects of the longitudinal sub-model as a “Fixed and Random effects” association structure, whereas those that only involve random effect but no fixed effects are designated “Random effects only”. The association parameter for association structures that involve both fixed and random effects quantifies the effect of the recorded longitudinal value for an individual on their risk of an event. Alternatively the association parameter for association structures that involve only random effects quantifies the effect of an individual’s deviation away from the population mean on their risk of an event. As such these association structures have different interpretations. A description of fixed and random effects association structures is given in Rizopoulos [52], whilst Henderson et al [30] discuss random effects only association structures. Additionally Rizopoulos and Ghosh [164] and Gould et al [75] discuss a range of association structures.

In total 33 studies (50.8%) reported at least one joint model that used a fixed and random effect association structure. A range of fixed and random effect association structures was reported (see Table 2). The current value association structure (24 studies, 36.9%) refers to models that inserted the value of the longitudinal trajectory at the current time point for the individual into their time-to-event sub-model, and is used when the effect of the current value of the longitudinal trajectory on the risk of an event is of interest. The current slope or first derivative of the fixed and random effects (3 studies, 4.6%) is when the first derivative of the longitudinal trajectory at the current time point is calculated. This association structure is used if the effect of the rate of change of the value of the longitudinal trajectory of an individual on the risk of an event is of interest. The current value and the current slope can be inserted into the model in tandem (5 studies, 7.7%), if both the effect of the current value and its current rate of change of the longitudinal trajectory are of interest. Finally 1 study (1.5%) reported a fixed and random effect association structure that inserted the coefficients of the fixed and random effects but not their covariates into the time-to-event sub-model.

Overall, 27 studies (41.5%) reported at least one joint model that used a random effects only association structure. The random effects only association structures also fall into distinct groups. Here, a random effects only association structure is defined to contain covariates if it is of a format such as  $(b_0 + b_1t)$  or  $\alpha_1b_0 + \alpha_2b_1 + \alpha_3(b_0 + b_1t)$  where the



$\alpha$  terms are association parameters, the  $b$  terms are random effects, and  $t$  represents a covariate such as time. Alternatively, the random effects only association structure is defined to not contain covariates if no covariates are present in the association structure, and it takes a format similar to  $\alpha(b_0 + b_1)$ , where the random effect  $b_1$  had a covariate  $t$  in the longitudinal sub-model (see Henderson et al [30] for further examples). Of the identified studies, 5 (7.7%) shared only a random intercept between the longitudinal and time-to-event sub-models. A total of 7 studies (10.8%) shared random effects with covariates, whilst 9 (13.8%) state that they shared only random effects, but without covariates, and in 6 studies (9.2%) it was unclear if the random effects were shared with or without their covariates.

Four studies (6.2%) were classed as falling into a specialist association structure group. This group contained less common association structures, such as linking the time-to-event and longitudinal sub-models through a multivariate normal distribution. Another notable group was the latent class structure [76, 162], used in at least one joint model in 3 studies (4.6%). Latent class joint models are a different joint model formulation to those that use the above association structures (which are termed shared parameter joint models). This thesis focuses only on the shared parameter joint models.

Finally there were 4 studies (6.2%), including the two studies only available as abstracts, for which the association structures were unclear.

### 2.3.1.9 Software

The software and packages used to fit joint models are listed in Table 3. Packages or methods have been stated even if no studies identified in the review currently used them.

The software and packages used to fit the joint model can often give information about the structure of the joint model. For example the *joiner* package in R only allows random effects only association structures, whilst the *JM* package in R only allows fixed and random effect association structures. However, the software used was not always stated by the identified studies, a potential issue for future MA when attempting to determine the structure of the joint model used.

The most commonly mentioned software was R, although SAS and Stata were also commonly stated. A possible reason for the preference of these software could be the availability of dedicated joint modelling packages in each. Some of the other software mentioned in the identified studies could require more coding input from the user (such as

C++ or Winbugs). Such software was used less frequently in the identified studies than the software with dedicated joint modelling packages, again adding credibility to the suggestion that software was picked for ease of use.

Several packages in R currently provide options for joint models. The current preference indicated by this review is for the *JM* package [81], which implements frequentist joint models that insert both the fixed and random effects of the longitudinal trajectory into the time-to-event sub-model. This preference could be attributable to the wide range of options available in the package, including range of baseline hazards and random effects specifications. Additionally, Rizopoulos' joint modelling textbook [52] provides detailed examples of implementation and interpretation of joint models using the package. Also of note is that the majority of analyses using SAS (10 of the 13 studies that fitted models using SAS) used the *PROC NLMIXED* function, although dedicated joint modelling macros for the software are now available.

In four studies (6.2%) more than one software was stated, it was unclear which implemented the joint model fit, and so these studies were grouped into a separate category.

Software	N (%)
<b>R [79]</b>	<b>21 (32.3)</b>
R ( <i>JM</i> ) [81]	15 (23.1)
R ( <i>JM</i> Bayes) [82]	0 (0)
R ( <i>joiner</i> ) [80]	1 (1.5)
R ( <i>frailtypack</i> ) [165]	0 (0)
R ( <i>JM</i> and <i>joiner</i> ) [80, 81]	1 (1.5)
R (unspecified package)	2 (3.1)
R (own code developed, unclear if available)	2 (3.1)
<b>SAS [85]</b>	<b>13 (20.0)</b>
SAS ( <i>PROC NLMIXED</i> )	10 (15.4)
SAS (own code available)	1 (1.5)
SAS (unspecified)	2 (3.1)
<i>JM</i> Macro [166]	0 (0)
<i>JMFit</i> Macro [167]	0 (0)
<b>Stata [84]</b>	<b>5 (7.7)</b>
Stata ( <i>stjm</i> ) [83]	2 (3.1)
Stata (unspecified)	3 (4.6)
<b>WinBUGS [168]</b>	<b>4 (6.2)</b>
WinBUGS (own code available)	2 (3.1)
WinBUGS (no available code)	1 (1.5)
WinBUGS (unspecified)	1 (1.5)
<b>OpenBUGS (no available code)</b>	<b>1 (1.5)</b>
<b>Fortran</b>	<b>3 (4.6)</b>
Fortran (code available, not study specific)	1 (1.5)
Fortran (own code developed)	1 (1.5)
Fortran (study states code available)	1 (1.5)
<b>NONMEM (unspecified) [169]</b>	<b>2 (3.1)</b>
<b>C++ (own code unclear if available) [170]</b>	<b>1 (1.5)</b>
<b>Mplus (unspecified) [171]</b>	<b>1 (1.5)</b>
<b>More than one software listed/potentially used</b>	<b>4 (6.2)</b>
R ( <i>JM</i> ) or SAS (unspecified)	1 (1.5)
R or SAS (unspecified)	1 (1.5)
WinBUGS and R (Directed Acyclic Graph provided)	1 (1.5)
WinBUGS and R (own code available)	1 (1.5)
<b>Unclear</b>	<b>10 (15.4)</b>

Table 3: Software used in joint model fits in included studies (note that studies could report multiple joint fits using different software, so total  $N \geq 65$ ). For McIlain [128] R code is stated as available in supplementary material, which was missing when accessed. Lawson [122] may have used WinBUGS but without seeking confirmation from the authors this was classed as unclear software. For Fortran see <http://www.fortran.com/>, accessed 10th February 2016)

### 2.3.2 Did studies report sufficient information to contribute to AD-MA?

The main aim of this review was to assess to what extent the current reporting practice of joint models for longitudinal and time-to-event medical or biostatistical data would allow AD-MA to be performed. As noted earlier, the feasibility of meta-analyses are considered separately for three separate groups: for the longitudinal sub-model parameters, for the time-to-event sub-model parameters and for the association parameters. For each of these groups, for an identified study to contain enough information for a MA to be possible, it must report a sample size and the relevant parameter estimate along with a precision estimate (which could be a standard error, or a confidence interval with a specified significance level). A summary of the proportions of studies reporting this information is given in Table 4, along with, for each of the three groups of interest, the proportion of MA possible given the information currently reported in the literature.

	Longitudinal MA	Time-to- event MA	Association MA
Coefficients reported (%)	45 (69.2)	46 (70.8)	51 (78.5)
Precision reported (%)	44 (67.7)	45 (69.2)	50 (76.9)
Standard Errors reported (%)	22 (33.8)	23 (35.4)	25 (38.5)
Confidence Intervals (CI) reported (%)	30 (46.2)	32 (49.2)	36 (55.4)
Significance level reported (%)	57 (87.7)	57 (87.7)	57 (87.7)
Sample size reported (%)	64 (98.5)	64 (98.5)	64 (98.5)
<b>MA possible given reported information (%)</b>			
All identified studies (N=65)	44 (67.7)	45 (69.2)	50 (76.9)
Studies using joint models to account for dropout (N=22)	18 (81.8)	14 (63.6)	15 (68.2)
Studies using joint models to include time varying covariate in time-to-event sub-model (N=4)	2 (50.0)	3 (75.0)	3 (75.0)

Table 4: Summary of information available to contribute to meta-analysis

The sample size was reported in the majority (98.5%) of the identified studies, with median sample size of 514 (IQR 277 to 1054.5, range 46 to 3814).

The association parameters were more commonly reported (51 studies (78.5%)) than those of the longitudinal and time-to-event sub-models (45 (69.2%) and 46 (70.8%) respectively). This could be attributable to the high proportion of identified studies that stated that joint models were used to investigate the link between longitudinal and time-to-event outcomes, which is quantified by the association parameter (Table 1).

The number of studies where a precision measure was available was comparable to the number of studies that reported coefficients, across the three MA categories. Whilst a MA

would be possible for each category if the parameter, the standard error and the sample size were reported, if the only available precision estimate was the confidence interval then the significance level was also required. In the studies identified, the significance level was unclear for 8 (12.3%) studies, was 0.05 for 53 (81.5%) studies, 0.01 for 3 (4.6%) studies, 0.1 for 1 (1.5%) study.

Overall, a MA would be possible for the association parameter group in 50 (76.9%) studies, for the longitudinal parameter group in 44 (67.7%) studies, and for time-to-event parameter group in 45 (69.2%) studies.

Ideally, enough information would be available to perform MA in all three groups of interest. Sufficient information to allow MA to be undertaken in all three groups was available from 38 (58.5%) of the studies. MA were possible in only two groups in 6 studies (9.2%), and only 1 in 13 studies (20.0%). In 8 studies (12.3%) there was insufficient information to complete any MA.

As noted earlier, the reasons for joint model use could affect what information is reported. The feasibility of each of the MA groups was re-examined dependent on the reasons stated for joint model use (Table 1).

For the 22 studies which stated “accounting for dropout” as a reason for joint model use, there was a higher percentage (81.8%) for which MA of longitudinal sub-model parameters was possible, compared to for all studies (67.7%). However, the percentage of MA that were possible for the time-to-event coefficients or association parameters was smaller. This could be explained by studies using joint models to account for dropout being mainly interested in the parameters from the longitudinal sub-model.

Only 4 studies stated inclusion of a time varying covariate in a time-to-event model as one of their reasons for using joint models. There was a slight indication that the longitudinal coefficients for studies using joint models to include time varying covariates in time-to-event models are worse reported than for all identified studies, possibly because the longitudinal component of the joint model is of interest as a covariate rather than an outcome in these cases, however more information is needed before this relationship can be fully investigated.

## 2.4 Discussion

Joint models for longitudinal and time-to-event data are often stated as beneficial compared to separate longitudinal or time-to-event analyses, as they can reduce bias and increase efficiency in model estimation (see Ibrahim et al [87] for example). Additionally, Powney et al [172] discuss a study where joint models showed a significant difference between treatment groups that was not identified by separate analyses [173, 174]. These benefits of joint models reinforce the suggestion that in certain circumstances MA of joint models may be more appropriate than MA of separate models. The review conducted in this chapter assessed the current use of joint modelling methods in single study analyses, with a view to their use in future meta-analyses.

### 2.4.1 Recommendations for reporting of future single study joint analyses

This review highlighted that single study joint analyses are currently implemented in a range of clinical areas, using a variety of modelling structures, implemented with an assortment of software. Whilst reporting of single study joint analyses was good, steps could be taken to ensure the information necessary to conduct a meta-analysis is included in publications. These recommendations are discussed below, and summarised in Figure 3.

The search strategy aimed to identify all studies that implemented joint models to influence future healthcare, and it is believed that the studies identified are representative of the current literature. However, if studies did not state the key search terms used in this review (see Appendix 1 for search strategies) in text accessible to the search, the study may not have been identified. For example, from Powney et al [172] it is known that the MAGNETIC trial [173, 174] utilised joint models, however this is not mentioned in the abstract. When joint models are not used as part of the primary analyses, their use may be unclear from the abstract or keywords. Therefore this research suggests that statement of statistical methods should be made in text accessible to search engines (see Figure 3).

With the increasing use of joint models in the literature, ensuring they are well reported is vital so that the analyses can be interpreted fully and that the published data can be used in future evidence synthesis. For the scenario of AD-MA of the results of joint models published in the literature, it was determined that it was possible to perform MA for a high proportion of, but not all, studies. For future practice it is recommended that regardless of the reason for joint model use, full model covariates with precision estimates be reported either in the study report or supplementary materials (Figure 3) not only to aid interpretation within the study itself, but to ensure that future MA would be possible.

Furthermore, it is important that model structure (both sub-models and the association structures that link them) are clearly reported (Figure 3). In some studies identified in this review, model formulae were not reported (a particular issue for association parameters). A wide range of association structures are available to researchers. Due to the different interpretation of each, it may not make sense to pool association parameters from radically different association structures (see Section 3.2.4). Without clear statement of the model structure it may be difficult to reliably conduct future MA. It would be beneficial if studies applying joint models included statement of the model structure as standard to give clarity to the methods used.

Additionally, details of the demographic of the population should be reported, as results could differ between different study populations. For example, a treatment may have a different effect in an older compared to a younger population. It is generally accepted in meta-analyses that if populations are too dissimilar, information from them should not be pooled. If populations are comparable, but differ in certain demographic characteristics that could affect the analysis (e.g. treatment effect differing across age groups), sub-group analyses can be performed in AD-MA (grouping studies included in the meta-analysis, e.g. by a categorical variable, and pooling effects within sub-groups). Whilst this may show differing treatment effect between different sub-populations, sub-grouping requires simplifications of data, such as categorisation of continuous treatment effect modifiers (such as age  $\leq 50$ , age  $> 50$ ). As such, if it is thought that treatment is affected in a complex way by covariates, or that detail more than that supplied by basic sub-group analyses in an AD-MA is needed, IPD may have to be sought, and analyses conducted that include the demographic covariates that could affect treatment (see Section 3.4.3). This is not an issue specific to the meta-analysis of joint data, and so has not been included in our recommendations for reporting of joint models, as it is good practice for any analysis of data to give description of the demographics of the study population.

In analyses of time-to-event data, the baseline hazard is not commonly reported. However it can communicate useful information about the progression of an illness [175].

Additionally, comparison of baseline hazards between study populations can indicate potential heterogeneity in the data, for example differences in populations may cause the disease to progress at different rates. Baseline hazards that differ significantly between studies (such as one study reporting a constant baseline hazard, and another reporting non-linear, non-constant baseline hazard) should motivate the researcher to examine whether it is appropriate to pool results from the two studies (this can be assessed ideally

through provided estimates of the baseline hazard, or if this is not available a comparison of chosen modelling method for the baseline hazard could be made). As such, it is beneficial if the baseline hazard estimates are reported for a joint modelling analysis, so that the shape of the baseline hazard can be viewed and compared between studies.

However, as baseline hazard is often viewed as a nuisance parameter, provided it has been adequately modelled, non-reporting of baseline hazard estimates would not exempt a study from contributing to a meta-analysis (resulting in reporting of baseline hazard not being included in recommendations for reporting of joint models, which aim to describe the minimum information that should be reported to a joint model to be able to contribute to AD-MA). Analyses utilising different (but appropriate) methods to model the baseline hazard can be pooled. If methods to model baseline hazards differ between studies, and estimates of baseline hazards are not supplied, sensitivity analyses can be undertaken to assess whether MA conclusions differ dependent on the method used to model the baseline hazard.

As it is useful if baseline hazard estimates are reported, the code produced during this research (which can fit joint models to both single study and multi-study data) returns estimates of the baseline hazard (see Section 4.3.2.2). A future extension to the package which may be beneficial are functions to plot the baseline hazard, and also alternatives to modelling the baseline hazard including spline approaches.

Also, it is important especially in joint modelling to state the software, packages or functions used to fit the joint models. In any investigation, statement of the software used to conduct the analysis is vital to ensure reproducibility and transparency of the research. However, the software used could indicate the joint model structure, as well as methods used to obtain the parameter estimations (Figure 3), as, for example, some packages only support certain types of association structures. Consequently, the software used to fit the model can also indirectly inform researchers as to the joint modelling methods used.

Consistently, the proportion of studies where association parameter MA was possible was higher than the proportion where MA was possible for longitudinal or time-to-event parameters. The association information may have been more commonly reported than the longitudinal or time-to-event information because the association parameters in shared random effect models quantify the link between the sub-models. It can be expected therefore that studies aiming to quantify the link between the sub-models report the association information more prominently than other model parameters.



#### **Recommendations for future reporting of joint models**

1. Model structure (longitudinal and time-to-event sub-models, and association structure) be clearly stated
2. Estimates of all model coefficients and their precisions be available in main or supplementary material
3. Software (and packages) used to fit models be stated, or code available on request from author
4. Statistical methods should be mentioned in text accessible to search engine to aid identification of papers in meta-analyses

*Figure 3: Recommendations for future reporting of joint models*

#### **2.4.2 Concluding Remarks**

Overall, this chapter has reported the results of a review of current reporting of joint models in the literature. It has highlighted the importance of fully reporting the coefficients and precision estimates of applied joint models in publications, as well as a clear statement of the structure of the joint model used and the software employed. Whilst the review identified a limited number of studies that reported sufficient information to contribute to a meta-analysis, there was great variation in the disease area and joint modelling structures. It may not be currently possible to identify sufficient studies in an area to conduct a published data AD-MA of joint model fits. Nevertheless this review has indicated that joint model publications are increasing in number, and the increasing availability and flexibility of joint modelling software is likely to facilitate their application further.

This chapter prompted many point of discussion for the two-stage IPD-MA of joint data. The methods of two-stage IPD-MA are closely linked to those of AD-MA: the second stage in such a two-stage process is the same as an AD-MA. Points of interest included the suitability of pooling association parameters from different joint models, the interpretation of parameters from different model structures and the process of selecting the methods to use in a joint modelling analysis. These and other points are discussed in Section 3.2.

In the future, this investigation recommends that reporting standards of joint model fits should be maintained and improved on, following the recommendations given here and in Sudell et al 2016 [93], in order that published applied joint models can contribute to future AD-MA.

## Chapter 3 : Methods for Meta-Analysis of Joint Longitudinal and Time-to-Event Data

---

In this chapter, I propose a range of methods to meta-analyse joint data generated by multiple studies. The chapter commences with a description of the structure of the data assumed available. Two-stage methods for the meta-analysis of multi-study joint data are then presented, and discussed. One-stage methods are then described, including a range of model groups each of which account for between study heterogeneity in different ways. The model groups are discussed and evaluated in turn. The chapter concludes with a discussion of meta-analytic joint modelling methodology.

### 3.1 Description of assumed data type

Throughout this chapter it is assumed that Individual Patient or Participant Data (IPD) is available. The methods described assume constant treatment effect over time and that time is reported as a continuous measure, measured from the same time point (e.g. time from randomisation into study) for each individual. Exact times at which longitudinal measurements are taken and the number of recorded measurements can differ between individuals. Only univariate joint models (those involving a single time-to-event and a single longitudinal outcome) are considered.

#### 3.1.1 General notation

The multi-study joint longitudinal and time-to-event IPD is considered to have three nested levels: longitudinal measurements at level 1, nested within individuals at level 2, nested within studies at level 3.

Study membership is identified by  $k = 1 \dots K$  where  $K$  represents the total number of studies in the meta-dataset.

Individuals within each study are represented by  $i = 1 \dots n_k$  where  $n_k$  denotes the total number of individuals in study  $k$ .

The longitudinal time points are identified using  $j = 1 \dots m_{ki}$  where  $m_{ki}$  represents the total number of longitudinal measurements recorded for individual  $i$  in study  $k$ .

The longitudinal outcome for individual  $i$  in study  $k$  at the  $j$ th time point is represented by  $Y_{kij}$ , whilst the survival time for individual  $i$  in study  $k$  is denoted by  $T_{ski}$  (where  $T_{ski}$  is the minimum of the individual's event time  $T_{Eki}$  and their censoring time  $T_{Cki}$ ).

Fixed effects are represented using  $\beta$  terms, with the first element of the subscript identifying the sub-model they belong to (such that  $\beta_1 = \beta_{11}, \beta_{12}, \beta_{13}, \dots$  are the longitudinal fixed effects, and  $\beta_2 = \beta_{21}, \beta_{22}, \beta_{23}, \dots$  are the time-to-event fixed effects).

Random effects are represented by  $b$ , with the level at which the random effects act identifiable through the bracketed superscript, such that individual level (level 2) random effects are represented by  $b_{ki}^{(2)}$  and study level (level 3) random effects by  $b_k^{(3)}$ .

Design matrices (matrices containing the covariates included in the model) are represented by  $X$  for the fixed effects and  $Z$  for the random. As with the fixed effect coefficients,  $X_1$  represents the longitudinal fixed effects design matrix, and  $X_2$  represents the time-to-event fixed effects design matrix. Additionally,  $Z_{ki}^{(2)}$  represents the design matrix for the individual level (level 2) random effects, and  $Z_k^{(3)}$  represents the design matrix for the study level (level 3) random effects.

Association parameters are represented by  $\alpha$ , with association parameters linked to individual level (level 2) random effects labelled  $\alpha^{(2)}$ , and those linked to study level (level 3) random effects as  $\alpha^{(3)}$ .

Finally, error terms are represented by  $\varepsilon$ . Additional notation will be defined as required. A table of notation used in the thesis is provided for reference at the start of the document on pages ix-x.

### 3.2 Methods for Two-stage IPD Meta-Analyses

In this section, methods for the two-stage meta-analysis of joint longitudinal and time-to-event IPD are discussed. The second stage of the methods could be used to perform an AD-MA of joint data, for example of published study level results, given that sufficient information was available to ensure that only parameters with comparable interpretations were pooled.

This section describes the process of a two-stage MA, starting with preliminary steps to assess the structure of the joint model, followed by a description of the first and second stages of the analysis. A set of guidelines is then presented, designed to ensure that

researchers conducting two-stage IPD-MA only pool parameters with comparable interpretations. The section concludes with a discussion of the two-stage methods.

The methodology covered in this section has been published as part of Sudell et al [176], including concise guidelines for the stages of a two-stage MA of joint data (presented here in detail).

### 3.2.1 Preliminaries

Before modelling the joint data available within each study, the most appropriate modelling structure must be selected. This includes choice of longitudinal and time-to-event sub-models, association structure, and a consideration as to the fixed and random effects to include. A variety of joint models are currently employed in the literature in analyses of medical data [93], see Chapter 2. Many choices exist for the longitudinal sub-model (including linear mixed effects models, or use of splines), or for the time-to-event sub-model (including PH or AFT models). To appropriately model the data, the most suitable family of sub-models for each of the longitudinal and time-to-event outcome must be selected. A useful method in this assessment is to produce plots of the study specific data.

Firstly, plots of the individual specific longitudinal trajectories panelled by event type (censored or experienced the event) should be generated. The time (displayed on the x-axis of the plot) can be the recorded longitudinal time-point, or can be adjusted by subtracting the survival time of each individual from their measurement times. This approach highlights changes in the longitudinal trajectories as time approaches the survival time. If the longitudinal trajectories show evidence of change (e.g. change in slope) immediately before survival time, there is evidence of an association or relationship between the longitudinal outcome and the time-to-event outcome.

Examination of the trajectories could indicate an appropriate random effects structure. For example, if variation in the trajectory intercept is apparent between individuals, an individual specific random intercept could be included in the study specific model. If the trend over time differs between individuals, an individual specific random time term may be beneficial. Through examination of these plots, the most appropriate sub-model family (random effects specification, linear or non-linear model) for the longitudinal outcome can be selected.

If the sample size within a study is large, inclusion of a “smoother” over the longitudinal trajectories can aid the assessment of the shape of trajectories (e.g. linear, non-linear, changes in trajectory shape before survival time). Another option is to colour each individual’s trajectory separately, allowing different individual’s longitudinal measurements over time to be distinguished in the plot containing all individuals within a study. Alternatively, samples of individuals from each study can be taken, and individual trajectories plotted side by side for each individual in trellis plots. This allows the behaviour of individual’s longitudinal trajectories to be clearly assessed.

If the measurement error in the longitudinal outcome is high, significant variability may be seen within the longitudinal measurements, even from the same individual. In this case, rather than producing trajectory plots that plot trajectory lines that join the longitudinal measurements recorded at each time point for the individual, a smoother can be produced for the measurements from each individual. This method would plot the general trend for an individual, allowing variation in measurements caused by larger measurement errors around this general trend.

Secondly the time-to-event data should be examined, to enable the family of time-to-event sub-models to be selected. This can be achieved by plotting the Kaplan-Meier [177] curves for each study. Examination of these curves, stratified by covariates of interest (for example treatment assignment), would help the most appropriate sub-model family for the time-to-event data to be determined (for example a PH or AFT model).

Additionally a choice for the modelling of the baseline hazard should be made. Options include flexible methods (including unspecified, piecewise constant or spline based) or parametric (including exponential or Weibull). As the baseline hazard itself is not of interest, just ensuring that it has been modelled appropriately, choice of methods can differ between studies. However, serious discrepancies between studies (such as one study having a constant risk of event over time, and another having a non-linear non-constant baseline hazard) should be motivation for further examination of the studies, to ensure that the study populations are comparable, and the results from the two studies appropriate to pool.

Finally, the association structure linking the sub-models must be selected. This is often chosen based on the aims of the investigation, and the clinical background of the data. Each association structure has a different interpretation [52, 75], and pooling parameters from different types of structure might be inappropriate (Section 3.2.4). Care should be

taken to select an association structure that has a meaningful interpretation to the investigation.

As the longitudinal and time-to-event outcomes are the same across the studies, the general behaviour of each outcome (e.g. nonlinear longitudinal trajectories), and so the most appropriate family of sub-models for each of the longitudinal and the time-to-event outcomes is likely to be similar across studies. If this is not the case, the reason for such disparity between studies in behaviour of outcomes should be investigated.

If there is uncertainty about the most appropriate model specification to employ in a joint modelling analysis, for example if there is uncertainty what random effects structure to use, multiple joint models representing differing potential specifications can be fitted. Models including different parameters can be compared, for example information criteria could be used [52, 77] such as Akaike's Information Criterion (AIC) [178], or the Bayesian Information Criterion (BIC) [179]. For these values, the smaller values of the AIC/BIC identify the better model. Alternatively, procedures such as forward or backward variable selection can be employed. In this way, the different potential model specifications suggested by the study specific plots can be compared, and an appropriate model specification selected. In practice, the choice of sub-models and association structures may be somewhat dictated by the availability of software to fit the required model, unless the researcher is willing to write their own code. For example, the two main frequentist univariate joint modelling packages in R (*joiner* [80] and *JM* [81]) permit different types of models. The *joiner* package allows random effects only association structures, with a Cox PH time-to-event sub-model with unspecified baseline hazard, and a linear mixed effects longitudinal sub-model. The *JM* package allows association structures involving both the fixed and random effects of the longitudinal sub-model (namely the current value or current slope structures). A variety of baseline hazards for the time-to-event sub-model are permitted, and either a parametric, a PH or an AFT model can be selected. The longitudinal sub-model can be a linear mixed effects model or can also contain splines. The *joinermeta* package (discussed further in Chapter 4) extends the *joiner* package to the multi-study data case, however is still restricted to the same association structures and model types as the *joiner* package. Other modelling options also exist in other software, for example the *stjm* [83] function in Stata. Consequently, researchers are faced with a choice between one of the modelling options currently available in pre-existing packages, or to invest time in writing code specifically for their project.

### 3.2.2 Stage 1 – Modelling of joint data within studies

In the first stage of a two-stage IPD-MA of joint data, joint models would be fitted to the IPD within each study. Studies identified in the meta-analysis should be grouped so that the chosen model structure within each group (sub-model types, association structure, and specification of any terms fixed or random that are involved in the association structure) are identical. Models of the same structure should then be fitted to each of the study specific datasets within each group. The recommendation of grouping studies, and fitting identical models within each group, is discussed in Section 3.2.4.

During this investigation of the two-stage MA of joint data, only joint models of the following format are considered. The longitudinal sub-model is a linear mixed effect model of specification:

$$Y_{kij} = \mathbf{X}_{1kij}\boldsymbol{\beta}_{1k} + \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)} + \varepsilon_{kij} \quad (13)$$

In equation (13), the longitudinal outcome is represented by  $Y_{kij}$ , and the design matrix for the fixed effects (containing the covariate values for individual  $i$  in study  $k$  at time point  $j$ ) is represented by  $\mathbf{X}_{1kij}$  with corresponding population coefficients  $\boldsymbol{\beta}_{1k}$ , which are estimated for each separate study  $k$ . The individual level random effects are represented by  $\mathbf{b}_{ki}^{(2)}$ , with corresponding design matrix  $\mathbf{Z}_{kij}^{(2)}$ . The random effects are assumed to follow a zero mean multivariate normal distribution  $N(0, \mathbf{D}_k)$ . The error term is represented by  $\varepsilon_{kij}$ , and is assumed to follow a  $N(0, \sigma_k^2)$  distribution, with error terms at each time point identically but independently distributed (IID). Within and across studies, the error terms and the random effects are considered independent of each other.

The time-to-event sub-model of the joint model assumes the specification in equation (14).

$$\lambda_{ki}(t) = \lambda_0(t)\exp(\mathbf{X}_{2ki}\boldsymbol{\beta}_{2k} + W_{2ki}(t)) \quad (14)$$

In equation (14),  $\lambda_0(t)$  represents an unspecified baseline hazard,  $\mathbf{X}_{2ki}$  represents fixed time stationary population covariates with associated coefficients  $\boldsymbol{\beta}_{2k}$ , which again are estimated for each separate study.  $W_{2ki}(t)$  is the association structure linking the time-to-event and longitudinal sub-models. A shared zero mean random effects proportional association structure is assumed, specifically,

$$W_{2ki}(t) = \alpha_k^{(2)}W_{1ki}(t) \quad (15)$$

Here  $W_{1ki}(t) = \mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}$  are the zero mean random effects of the longitudinal sub-model, and  $\alpha_k^{(2)}$  quantifies the association between the longitudinal and the time-to-event outcomes. This association parameter can be interpreted as the effect of the individual deviation from the population mean trajectory value at a given time point on the individual's risk of an event.

### 3.2.3 Stage 2 – Pooling of results between studies

In the second stage, estimates of parameters of interest are extracted from the study specific model fits and pooled using standard meta-analytic techniques. During this thesis, both fixed and random effect MA are conducted and the results compared. The Inverse Variance approach is employed in both cases (using the DerSimonian and Laird [10] approach for random MA, see Whitehead [2] for a MA overview).

A fixed effect MA assumes that each study is estimating a common underlying effect denoted by  $\theta$ . Any variability between studies is attributed to sampling variability. Study specific estimates are then pooled using a weighted average:

$$\hat{\theta} = \frac{\sum_{k=1}^K w_k \hat{\theta}_k}{\sum_{k=1}^K w_k} \quad (16)$$

In equation (16),  $\hat{\theta}_k$  is the effect estimate from the  $k$ th study,  $\hat{\theta}$  is the overall pooled effect estimate, and  $w_k$  is the weight for study  $k$ . For the inverse variance method this specified as  $w_k = 1/v_k$  where  $v_k = \text{var}(\hat{\theta}_k)$ .

A random effects MA assumes that the observed study specific effect estimates are a random sample from a distribution of possible effect estimates, and pooled results estimate the mean effect and its variance for the population. If the between study variability in effect estimates is represented by  $\tau^2$ , then the variance of the effect estimate is  $\text{var}(\hat{\theta}_k) = v_k + \tau^2$ . Results are then pooled using equation (16), but with the weight term now equal to  $w_k = 1/(v_k + \tau^2)$ .

As mentioned, parameters should only be pooled within groups, not across groups. This ensures that the parameters have comparable interpretations, as joint models fitted to studies within the same group employ the same model structure (sub-model types, association structure, and specification of any terms fixed or random that are involved in the association structure). As stated, the methodology employed in this chapter assumes availability of IPD, allowing similar models (provided they are appropriate) to be fitted to the data from each study. However, the second stage of this process can be applied to AD-



MA. In this case, it may be anticipated (given the variability in joint model structure observed in Chapter 2) that there are not many analyses that fall within the same group. If this is the case, in order to ensure comparable interpretation of pooled parameters, it is recommended to seek IPD from each study included in the meta-analysis.

### 3.2.4 Recommendations for a two-stage IPD-MA of joint data

Whilst this thesis is concerned only with joint models that employ an association structure that shares zero mean random effects with a common association parameter, a variety of random effect and association specifications exist for joint models [75]. However, dependent on the model specified, association parameters can have very different interpretations.

For example, consider a joint model that employs an association structure that shares only zero mean random effects between the longitudinal and time-to-event sub-models with a common association parameter (such that  $W_{2ki}(t) = \alpha_k^{(2)} (\mathbf{Z}_{kij}^{(2)} \mathbf{b}_{ki}^{(2)})$ ). Here,  $\alpha_k^{(2)}$  quantifies how the deviation of individual  $i$  in study  $k$  from the population mean longitudinal trajectory at a given time affects their risk of an event [75]. An example could be a case where individuals with higher blood pressure than the population average at a given time are at higher risk of a cardiac event.

Alternatively, a joint model that employs an association structure involving the true complete longitudinal trajectory with a common association parameter (such that  $W_{2ki}(t) = \alpha_{k\_current} (\mathbf{X}_{ki} \boldsymbol{\beta}_{1k} + \mathbf{Z}_{kij}^{(2)} \mathbf{b}_{ki}^{(2)})$ ) has a different interpretation. Here,  $\alpha_{k\_current}$  quantifies how the value of the longitudinal outcome for individual  $i$  in study  $k$  at a given time point affects their risk of an event [52, 75]. An example could be where the risk of a cardiac event for an individual increases as their recorded blood pressure increases.

The above two association parameters are modelling different types of links between the sub-models. Pooling the two association parameters described in the example above would assume that the risk of an event at a given time point due to the difference in the individual's recorded longitudinal value and the population mean, and the risk of an event due to individual's recorded longitudinal value, are the same.

The size of the difference in association parameter estimates between the current value and random effects only examples mentioned above (assuming the same random effects structure) depends on the scale of the longitudinal measurements, and their variability between individuals. Longitudinal outcomes that take large values on a scale (e.g. values of

several hundred) but show little variability between individuals, could result in the expression  $\mathbf{X}_{ki}\boldsymbol{\beta}_{1k} + \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  having larger values, whilst  $\mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  taking smaller values. Alternatively, longitudinal outcomes that take smaller values on a scale, but display clear variability between individuals could report more similar values of  $\mathbf{X}_{ki}\boldsymbol{\beta}_{1k} + \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  and  $\mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  for a given individual. The significance of the association parameter informs the researcher whether the association structure represents an existing relationship between the longitudinal and time-to-event outcomes. The effect of the association structure on risk of an event is found by the product of the association parameter and the sum of the shared terms. As such if the sum of the shared terms is large (e.g.  $\mathbf{X}_{ki}\boldsymbol{\beta}_{1k} + \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  is large), then the association parameter estimate might be small in magnitude, but significant. To summarise, the interpretation of the association structure is separated into whether the association structure represents a true relationship between the longitudinal and time-to-event outcomes (as judged by the significance of the association parameter) and the effect the association parameter has on risk of an event (as judged by the value of  $W_{2ki}(t)$ , the product of the association parameter and the sum of the shared terms).

If the specification of the longitudinal sub-model differs between joint models fitted, the terms linking the sub-models in the association structure may differ, again leading to association parameters with differing interpretations. For example, consider two joint models each linking the sub-models using  $W_{2ki}(t) = \alpha_k^{(2)} \left( \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)} \right)$ , one with random effects specifications of random intercept alone  $\left( \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)} = b_{0ki}^{(2)} \right)$  and the other employing a random intercept and slope  $\left( \mathbf{Z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)} = b_{0ki}^{(2)} + b_{1ki}^{(2)}\text{time}_{ki} \right)$ . In the first case, the  $\alpha_k^{(2)}$  parameter quantifies the difference in risk of an event due to the difference between the population and individual specific intercept of the longitudinal trajectory. In the second,  $\alpha_k^{(2)}$  quantifies the risk of an event due to differences between the individual and the population both in the intercept and the slope of the longitudinal trajectory.

As such, differences in interpretation of association parameters can stem from the type of association structure used, or differences in terms included in the joint model. Pooling association parameters with different interpretations assumes that these differences are unimportant, potentially a strong assumption. To avoid this, it is recommended to pool association parameters from joint models with identical association structures (family of association structure, and the terms it involves). If differing joint models are appropriate

between studies identified in a meta-analysis, studies could be sub-grouped, with joint models of identical specification within but not across sub-groups. This is reiterated in the recommendations for two-stage meta-analysis of joint data, see Sections 3.2.4.2-3.2.4.3.

In addition, a variety of options for the longitudinal and time-to-event sub-models have been employed in the literature [93]. Pooling fixed effect coefficients from different types of sub-model again assumes their interpretation is comparable, which may not always be the case (such as for coefficients from PH models and AFT models). Again, care should be taken when employing joint models in a two-stage analysis to only pool parameters with comparable interpretations. The recommendation of sub-grouping studies such that the specification of joint models applied to each study within each group ensures that only parameters with comparable interpretations are pooled.

To ensure only appropriate pooling of parameters from joint model fits, the following recommendations are proposed for researchers to consider when conducting a two-stage MA of joint longitudinal and time-to-event IPD.

#### **3.2.4.1 Preliminary Work**

- For each study in the meta-analysis produce the following plots:
  - Longitudinal trajectories plots panelled by event type such that the behaviour of trajectories of those experiencing an event and those censored is clearly visible
  - Kaplan Meier plots, with curves stratified by covariates of interest for example treatment assignment.
- Using these plots:
  - Assess if an association might be present between the longitudinal and time-to-event outcomes (indicated for example by differences in the longitudinal trajectories for those experiencing an event and those censored).
  - For each study, identify an appropriate longitudinal sub-model family. Additionally consider what types of random effects structures (random intercept, random slope (time) term etc.) are appropriate
  - For each study, identify an appropriate time-to-event sub-model family
- Determine the type of association structure that best suits the aims of the analysis e.g. if clinically the amount of deviation between an individual's and the population average longitudinal trajectory at a given time is thought to affect the risk of an

event, then a zero mean random effects only sharing structure should be employed. Alternatively, if there is clinical evidence to suggest that the rate of change of the longitudinal trajectory affects the risk of an event, then the first derivative of the longitudinal trajectory should be shared between sub-models.

#### 3.2.4.2 *First Stage*

- Group the studies identified in the meta-analysis so that the chosen model structure within each group (type of longitudinal sub-model, type of time-to-event sub-model, association structure, random effect specification, and if both fixed and random effects are shared between sub-models specification of fixed effects in the longitudinal sub-model) is identical.
- Within each sub-group, fit identical joint models to data from each study. Model structures can differ between groups.

#### 3.2.4.3 *Second Stage*

- For each study  $k$  extract parameters of interest (subsets of the fixed effects from the longitudinal sub-model  $\beta_{1k}$ , of the fixed effects from the time-to-event sub-model  $\beta_{2k}$  and the association parameter(s)  $\alpha_k^{(2)}$ ) from the model fit, along with estimates of their precisions, and the sample size of the study.
- Pool estimates within each group using the inverse variance methods described in Section 3.2.3. Perform MA separately for each longitudinal coefficient, time-to-event coefficient and association parameter of interest.
- Results between groups with different joint model specifications can be qualitatively compared in discussion of the meta-analysis.

### 3.3 Methods for One-stage IPD Meta-Analyses

In this section, methods for the one-stage meta-analysis of joint longitudinal and time-to-event IPD are discussed. The clustering of data within studies must be accounted for in some way in the model. A variety of methods exist to accomplish this, including fixed interaction terms between covariates of interest and the study membership variable, study level random effects, or baseline hazard stratified by study.

Whilst joint models accounting for clustering using fixed interaction terms with study membership variables were possible to implement at the start of this PhD using currently available software, no general software was available that designed specifically to model multi-study joint data. Some papers fitted multi-centre joint models (e.g. Brombin et al

[90]), however they did not provide generalisable code. As such, as part of this thesis, I aimed to develop an R package *joinerMeta* to allow joint models to be fitted to multi-study joint data accounting for clustering using either fixed interaction terms, using study level random effects, stratifying baseline hazard or by using a combination of these methods. Although not the focus of this thesis, it is possible to model multi-centre joint data using this software. Descriptions of the software produced are available in Chapter 4, while theory behind the model fitting performed by the package is discussed in Chapter 5.

### 3.3.1 Preliminaries

The notation already defined in this chapter is adopted, but extended to account for study level as well as individual level random effects.

As the models investigated in this section have exact specifications of included fixed and random effects, individual covariates are identified by name rather than labelling  $x_1, x_2, \dots$ . Covariates additional to those mentioned can be included in the models fitted by the *joinerMeta* package.

During the discussion of one-stage methods, again the investigation is restricted to proportional random effects only association structures. However, some of the models considered allow random effects both at the individual and at the study level, with separate association parameters for each level of random effects. Separate association parameters are permitted for each level of random effects as the level of association between the longitudinal and the time-to-event outcomes may differ between the individual and the study level.

Consider a case where higher values of the longitudinal outcome results in increased risk of an event. At an individual level, an individual who reports longitudinal outcome larger than the study population average at a given time point would have an increased risk of an event. But at a study level, a study that has a higher study level average longitudinal outcome at a given time point than other studies, will contain a population of individuals at higher risk of an event. The differences between study populations may be much greater than the differences between individuals within a study. As such, distinguishing between association based on individual level and on study level random effects was appropriate.

As a real example, imagine a joint model with individual level random intercept and slope, and study level random intercept and treatment effect applied to data with longitudinally measured blood pressure and time until death. If the association parameters for both levels

are positive and significant, then study populations with average study blood pressure higher than the overall population average at a given time are at higher risk of death. In addition, individuals within a study with recorded blood pressure higher than the study mean at a given time are at higher risk of death.

In some cases, it may not be appropriate to separate the association parameter in this way. A future area of research beyond this thesis is to extend the software developed to allow for a range of association structures, such as those discussed in Gould et al [75]. Future research aims are discussed fully in Chapter 8.

### 3.3.2 Methods to account for between study heterogeneity

As noted earlier, the clustering of data within studies must be accounted for in the model. If differences exist between the data stemming from different studies, termed between study heterogeneity, this can be estimated or accounted for in the model. Various options exist, including fixed interaction terms between covariates thought to have differing effects between studies and study membership, study level random effects, or (in the case of the time-to-event sub-model) stratification of the baseline hazard by study. As discussed in Chapter 1, a considerable amount of work exists for separate IPD one-stage meta-analysis of longitudinal outcomes [31] and time-to-event outcomes [56, 180]. Drawing on the existing literature for these two areas, during this section a selection of potential modelling structures for one-stage meta-analytic models for joint longitudinal and time-to-event data containing a single longitudinal and a single time-to-event outcome are explored.

Six main groups of models are examined, detailed in Table 5. The models presented are restricted to involve only longitudinal time, treatment assignment and study membership as covariates of interest. Each of the 6 groups accounts for between study heterogeneity in a different way. Comparisons of the different methods will be made, and their advantages and disadvantages discussed.

If variables other than study membership, treatment assignment and longitudinal time are of interest to the investigation, they could also be included in the model in various formats (i.e. with interaction terms or with study level random effects), making the one-stage multi-study joint models a flexible method to investigate the effect of covariates on between study heterogeneity. However, during this thesis only the effect of treatment assignment on between study heterogeneity is examined. Software to enable the flexible modelling of multi-study joint data, allowing flexibility for future investigations, is discussed in Chapter 4.

The choice of what model group to use in a one-stage meta-analysis should be made on a case by case basis. For example, if it is important for the investigation to produce study specific estimates of the treatment effect for both the longitudinal and the time-to-event outcomes, then a model from group 1, which estimates interaction terms between the study membership and the treatment assignment covariates, should be chosen. However, if the meta-analysis contains a large number of studies, a model that accounts for between study heterogeneity using study level random effects (e.g. model groups 2, 3, or 5), may be more appropriate. Model groups are discussed and contrasted in the following sections.

As with the methods for two-stage analysis of multi-study joint data (Section 3.2), before fitting joint models to the data, it is recommended to produce study specific plots of the data. These plots have been described and discussed in Sections 3.2.1 and 3.2.4.1, and include study specific Kaplan-Meier plots and study specific longitudinal trajectory plots (plotting longitudinal outcome both against measurement times, and against measurement times minus each individual's survival time). Using these plots, as for the two-stage procedures, appropriate longitudinal sub-model and time-to-event sub-model can be examined. For example, the longitudinal trajectory plots could indicate that functions of time need to be included in the model, or the Kaplan-Meier plots could indicate that a proportional hazards assumption is reasonable. Comparisons of the plots between studies would also help to indicate the level of between study heterogeneity; if the plots between studies are similar, the level of between study heterogeneity is likely to be low.

Model Group	Longitudinal Sub-model	Time-to-event Sub-model	Association Structure
0	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
1	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + \beta_{14}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + \beta_{23}treat_{ki} * study_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
2	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$
3	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki})$
4	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + \beta_{14}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
5	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}time_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$

Table 5: Table of basic model structure of models investigated to account for between study heterogeneity in one-stage analyses. Longitudinal time time<sub>kij</sub> in the shared random effects is replaced with survival time T<sub>Ski</sub>.



### 3.3.2.1 Group 0: Naïve model

The first group of models (referred to as group 0, Table 5) is a naïve group of models, which does not account for between study heterogeneity in any way. It is not recommended to use such a model where heterogeneity exists between studies, and is presented in this investigation for comparison to the models which do account for differences between studies. This model represents the base model to which alterations are made to account for the multi-study nature of the data. In an analysis, between study heterogeneity would be tested for by comparing models from groups that accounted for between study heterogeneity in some way (e.g. through fixed or random effects, or stratification of baseline hazard by study) and this model group that ignores between study heterogeneity. This can be done for example by using information criteria such as the AIC/BIC [52, 77, 178, 179], where smaller values of the information criteria indicate better model fit. If the model accounting for between study heterogeneity gives a better fit, there is an indication that between study heterogeneity is present.

The longitudinal sub-model contains a population fixed intercept, longitudinal time ( $t_{kij}$ ), and treatment assignment ( $treat_{ki}$ ), as well as individual specific random intercept ( $b_{0ki}^{(2)}$ ) and time ( $b_{1ki}^{(2)}$ ) terms and an error term ( $\varepsilon_{kij}$ ). The error terms are independently and identically distributed at each time point following  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ . The random effects follow a zero mean multivariate normal distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ , where  $\mathbf{D}$  is a 2 by 2 covariance matrix with on diagonals giving the variance of the random effects and the off-diagonal giving the covariance between the random effects.

The time-to-event sub-model contains a fixed treatment assignment term ( $treat_{ki}$ ), and the association structure term  $W_{2ki}(t) = \alpha^{(2)} \left( b_{0ki}^{(2)} + b_{1ki}^{(2)} T_{Ski} \right)$ , which involves the zero mean individual level random effects between the sub-models with common association parameter.

### 3.3.2.2 Group 1: Fixed interaction term with study membership variable

The group of models labelled group 1 (Table 5) accounts for between study heterogeneity through the inclusion of the study membership covariate, and the interaction between study membership and treatment assignment, as fixed or population effects.

The longitudinal sub-model contains a fixed intercept, time ( $t_{kij}$ ), treatment assignment ( $treat_{ki}$ ), study membership ( $study_{ki}$ ) and interaction between treatment assignment and study membership (with corresponding coefficients  $\beta_{10}$  to  $\beta_{14}$ ), as well as individual specific random intercept and slope (time) terms ( $b_{0ki}^{(2)}$  and  $b_{1ki}^{(2)}$ ), and an error term  $\varepsilon_{kij}$ . The random effects and the error term are considered to be independent. The individual level random effects follow distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ , where  $\mathbf{D}$  is a 2 by 2 covariance matrix, whilst the error terms each follow distribution  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ . The time-to-event sub-model contains treatment assignment, study membership and an interaction between treatment assignment and study membership as fixed effects (with corresponding coefficients  $\beta_{21}$  through  $\beta_{23}$ ), as well as an unspecified baseline hazard common across included studies. Again, the association structure shares the zero mean individual level random effects between the sub-models with common association parameter.

This model group yields study specific estimates for the treatment effect for both sub-models. These are calculated, for the longitudinal outcome, by the sum of the  $\beta_{12}$  coefficient and each study's  $\beta_{14}$  parameter (a  $\beta_{14}$  coefficient will be produced for each study apart from that considered to be the baseline study due to the factor being transformed into dummy variables, so if for example there are  $K$  studies, there will be  $K - 1$  unique study specific  $\beta_{14}$  parameters). The study's estimate of  $\beta_{14}$  quantifies the difference in treatment effect between a particular study and the study specified as the baseline study. For the time-to-event outcome, the study specific coefficient is calculated by the sum of the  $\beta_{21}$  and each study's  $\beta_{23}$  parameter, and this represents the study specific log hazard ratio of risk of an event between the treatment groups. Consequently, for this model group, exact estimates of the treatment effect in each study are available for each sub-model, and so the between study heterogeneity is not just accounted for, the exact differences between treatment effect in each study can be examined.

The model also includes fixed study membership terms in each sub-model. For the longitudinal sub-model, this results in study specific estimates of trajectory intercepts being obtained from the sum of the intercept  $\beta_{10}$  and the coefficient for the relevant study dummy variable  $\beta_{13}$  coefficient. The study's estimate of  $\beta_{13}$  quantifies the difference in intercept between a particular study and the baseline study. In the time-to-event sub-model, the  $\beta_{22}$  parameter

represents the difference in log hazard ratio of risk of an event between the study in question, and the baseline study.

A comparison of the deviance ( $-2 * \ell(\theta)$ , where  $\ell(\theta)$  is the log-likelihood) between a model including the fixed study membership and interaction terms, and one without can identify if between study heterogeneity exists. The difference in the deviances will follow a  $\chi^2$  distribution with  $P$  degrees of freedom, where  $P$  is the difference in number of parameters between the two models. This comparison can only be used for nested models.

There are important considerations for this model group. For example, the choice of baseline study is important as it should be representative of the overall population of interest. In addition, the treatment effect estimates from this group of models are study specific, whereas the aim of a meta-analysis is generally to provide an overall pooled parameter estimate for the entire population.

Also, this model assumes that the baseline hazard in each study is proportional to a common baseline hazard. This may not be a suitable assumption given the differing populations and designs of the studies included in the meta-analysis. Assuming proportional hazards within studies rather than across all studies can be achieved by stratifying the baseline hazard by study, as examined in model groups 4 and 5 (Sections 3.3.2.5 and 3.3.2.6).

Finally, as the interaction terms result in a parameter per study for treatment effect in both sub-models, as the number of studies in the meta-analysis increases, the number of parameters estimated will also increase. There may be a cut-off point over which the number of studies makes this model unwieldy. This issue is investigated during the simulation study presented in Chapter 7.

### 3.3.2.3 Group 2: Fixed study indicator, study level random treatment effect

The second group of models introduces study level random effects to the multi-study joint model (Table 5). This group contains a longitudinal sub-model with population fixed intercept, time ( $t_{kij}$ ), treatment assignment ( $treat_{ki}$ ), and study membership ( $study_{ki}$ ) terms, with coefficients  $\beta_{10}$  to  $\beta_{13}$ , along with two groups of random effects ( $\mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}$ ), and error term ( $\varepsilon_{kij}$ ).

The individual level (level 2) random effects ( $\mathbf{b}_{ki}^{(2)}$ ), as before, consist of an individual level random intercept ( $b_{0ki}^{(2)}$ ) and a random time term ( $b_{1ki}^{(2)}$ ), following distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ . The study level (level 3) random effects ( $\mathbf{b}_k^{(3)}$ ) consists of a study level random treatment effect ( $b_{1k}^{(3)}$ ) following normal distribution  $\mathbf{b}_k^{(3)} \sim N(0, \sigma_A^2)$ , where in this case  $\sigma_A^2$  is a single value giving the variance of the single study level random effect. The random effects acting at different levels are considered independent of each other, and independent of the error terms. The error terms at each time point are independently and identically distribution following  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ .

The time-to-event sub-model contains a population fixed treatment assignment ( $treat_{ki}$ ) and study membership terms ( $study_{ki}$ ) with coefficients  $\beta_{21}$  and  $\beta_{22}$ , with an unspecified baseline hazard common across the studies in the meta-analysis.

The association structure (given in Table 5) involves all random effects included in the model, with common association parameter  $\alpha^{(2)}$  for the shared individual level random intercept and slope, and common association parameter  $\alpha^{(3)}$  for the shared study level random treatment effect. Occurrences of the longitudinal time variable  $t_{kij}$  in the association structure are replaced with the individual's survival time  $T_{ski}$ .

In this group of models, between study heterogeneity is accounted for in two ways. Firstly, a fixed study membership variable is included in both sub-models. In the longitudinal sub-model this adjusts the intercept for the baseline study for the longitudinal trajectory  $\beta_{10}$  by a study specific amount quantified by the  $\beta_{13}$  parameters (one of which will be calculated for each study apart from the baseline study in the meta-analysis). So (apart from the baseline study) for study  $k$  for the longitudinal outcome there is a study specific intercept of  $\beta_{10} + \beta_{13k}$  where  $\beta_{13k}$  represents the coefficient for the dummy variable for membership to study  $k$ . Inclusion of the fixed study membership variable in the time-to-event sub-model adjusts the log hazard ratio of risk of an event in the baseline study by a set amount for each study, quantified through the coefficient for the dummy variable for membership to study  $k$ .

Between study heterogeneity in treatment effect is quantified using a random effect, through inclusion of the  $b_{1k}^{(3)}$  term. The interpretation of treatment effect is more complex in joint one-stage MA models for separate longitudinal or time-to-event one-stage MA models due to its

presence in both sub-models. In the longitudinal sub-model, the  $b_{1k}^{(3)}$  term adjusts the overall population treatment effect coefficient  $\beta_{12}$  to give the observed treatment effect in study  $k$  through expression  $\beta_{12} + b_{1k}^{(3)}$ . Through the association structure, the study level random treatment effect  $b_{1k}^{(3)}$  is also present in the time-to-event sub-model. As such, the population effect of treatment group  $\beta_{21}$  is altered to give a study specific estimate of the deviation in log hazard ratio for risk of an event due to treatment group given by  $\beta_{21} + \alpha^{(3)}b_{1k}^{(3)}$ .

As a result, between study heterogeneity is assessed in two ways. Between study heterogeneity due to treatment effect is assessed through the variance of the study level random effect  $\sigma_A^2$ . The presence of between study heterogeneity due to treatment effect can be assessed by comparing models with and without the study level random treatment effect using information criteria such as the AIC/BIC [52, 77, 178, 179], where smaller values of the information criteria indicate better model fit. If the model including the study level random treatment effect is preferred, then study level heterogeneity in the treatment is indicated, with large values of  $\sigma_A^2$  indicating greater levels of between study heterogeneity.

Once the model has converged, values of the potential study specific treatment effect parameter estimates can be obtained by generating realisations from normal distributions. Let the estimated distribution of the study level random treatment effect  $b_{1k}^{(3)}$  be  $N(0, \hat{\sigma}_A^2)$ . Therefore, to assess possible study specific longitudinal treatment effect estimates from hypothetical studies under this model, realisations from a  $N(\hat{\beta}_{12}, \hat{\sigma}_A^2)$  distribution must be generated. To assess possible study specific time-to-event treatment effect estimates under this model, realisations from a  $N(\hat{\beta}_{12}, (\hat{\alpha}^{(3)})^2 \hat{\sigma}_A^2)$  distribution must be generated. These realisations (their mean, range, spread etc.) can be visualised e.g. by producing a histogram of the realisations. The more spread out the values, the greater the between study heterogeneity.

Residual heterogeneity other than that in the treatment effect is tested for by comparison of the deviance of models with and without study membership included in the sub-models, which will follow a  $\chi^2$  distribution with  $P$  degrees of freedom, where  $P$  is the difference in number of parameters between the two models.

One advantage of group 2 models compared to group 1 models is that as the number of studies increases in the meta-analysis, the number of parameters to be estimated does not increase as severely (although there is still an increase due to the presence of the fixed study membership covariate in each sub-model). Contrastingly, the model includes a study level random treatment effect, whose distribution might be poorly estimated if there are few studies included in the meta-analysis. As such, for this model, a balance between enough studies to reliably estimate the study level random effects, but not so many that the model becomes unwieldy due to the fixed effects, may be difficult to obtain. The issue of number of contributing studies is investigated further through simulation study in Chapter 7.

As with group 1, this group produced study specific estimates of the longitudinal intercept, and of the log hazard ratio for each study. However, it does not provide fixed study specific estimates of the treatment effect. Rather, it adjusts for between study heterogeneity using study level random effects, and quantifies the level of between study heterogeneity through the estimate of the variance of the study level random effects. As meta-analyses aim to produce an overall pooled estimate of the parameter of interest (in this case treatment effect), not outputting study specific fixed estimates of treatment effect may not be an issue.

In addition, this model group employs a common baseline hazard across studies, and so still assumes that hazards are proportional across studies. As noted earlier, this may be an unreasonable assumption if the study demographics differ significantly.

#### 3.3.2.4 Group 3: Study level random intercept and treatment effect

The third group of models relies completely on study level random effects to model between study heterogeneity (Table 5). This group of models involves a longitudinal sub-model containing a fixed or population intercept, longitudinal time term ( $t_{kij}$ ) and treatment assignment covariate ( $treat_{ki}$ ), with coefficients  $\beta_{10}$  through  $\beta_{12}$ , as well as two sets of random effects ( $\mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}$ ) and an error term ( $\varepsilon_{kij}$ ).

The first set of random effects act at the individual level (level 2), and include an individual level random intercept ( $b_{0ki}^{(2)}$ ) and a random slope ( $b_{1ki}^{(2)}$ ). The individual level random effects follow distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ , where  $\mathbf{D}$  is a 2 by 2 covariance matrix. The second set act at the study level (level 3), and include a study level random intercept ( $b_{0k}^{(3)}$ ) and random

treatment effect ( $b_{1k}^{(3)}$ ). The study level random effects follow distribution  $\mathbf{b}_k^{(3)} \sim N(\mathbf{0}, \mathbf{A})$ , where  $\mathbf{A}$  is a 2 by 2 covariance matrix. For both  $\mathbf{A}$  and  $\mathbf{D}$ , the on-diagonals of the covariance matrix are the variances of the random effects at the respective levels, and the off-diagonals contain the covariance between the random effects at each level. The error terms are independently and identically distribution at each longitudinal time point, following  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ .

The time-to-event sub-model contains a fixed population treatment assignment covariate ( $treat_{ki}$ ) with coefficient  $\beta_{21}$ , with an unspecified baseline hazard common across studies. The association structure (see Table 5 for full expression) involves both the individual level and the study level random effects, with common association parameter  $\alpha^{(2)}$  for the shared individual level random effects, and common association parameter  $\alpha^{(3)}$  for the study level random effects.

As mentioned,  $\mathbf{A}$  is the covariance matrix for the study level random effects, and so quantifies the between study heterogeneity. As such the first on diagonal element of  $\mathbf{A}$  (termed  $\mathbf{A}_{11}$ ) is the variance of the study level random intercept, and quantifies the between study variability in the intercept. The second on-diagonal element  $\mathbf{A}_{22}$  is the variance of the study level random treatment effect, and quantifies the between study variability in treatment effect. The larger these values, the more variability between studies in the intercept and the treatment effect respectively.

The presence of between study heterogeneity in model group 3 can be assessed by comparing models with and without the study level random effects using information criteria such as the AIC/BIC [52, 77, 178, 179], where smaller values of the information criteria indicate better model fit. If the model including the study level random effects are preferred, then study level heterogeneity indicated, with large on-diagonal values of  $\mathbf{A}$  indicating greater levels of between study heterogeneity.

The effect of the adjustments of different parameters due to study level random effects can be examined again by producing realisations from the relevant normal distributions. Specifically, once the model has converged, potential hypothetical study parameter estimates can be generated from normal distributions based on the estimated model parameters. Let the estimated distribution of the study level random treatment effect  $\mathbf{b}_k^{(3)}$  be  $N(\mathbf{0}, \mathbf{A})$ . To assess

possible longitudinal intercept and treatment effect estimates for hypothetical studies under this model, generate realisations from:

$$N\left(\begin{pmatrix} \hat{\beta}_{10} \\ \hat{\beta}_{12} \end{pmatrix}, \hat{\mathbf{A}}\right)$$

To assess possible time-to-event intercept and treatment effect estimates for hypothetical studies under this model, generate realisations from:

$$N\left(\begin{pmatrix} 0 \\ \hat{\beta}_{12} \end{pmatrix}, (\hat{\alpha}^{(3)})^2 \hat{\mathbf{A}}\right)$$

Histograms of the realisations can help to visualise the mean, range and spread of potential study specific estimates from hypothetical generated studies; the more spread out the values, the greater the between study heterogeneity. The interpretation of the study level random effects in the one stage joint MA model is more complex than for separate one-stage longitudinal or time-to-event MA models. The study level random effects are present in both the longitudinal and the time-to-event sub-model due to their involvement in the association structure. As such, the study level random intercept  $b_{0k}^{(3)}$  causes the longitudinal intercept for study  $k$  to equal  $\beta_{10} + b_{0k}^{(3)}$ , but also  $\alpha^{(3)} b_{0k}^{(3)}$  represents the deviation in log hazard ratio for risk of an event in the  $k$ th trial from the population average taken across all studies in the meta-analysis.

The study level random treatment effect  $b_{1k}^{(3)}$  estimated for study  $k$  adjusts the population fixed treatment effect in the longitudinal sub-model  $\beta_{21}$  to give the treatment effect observed in study  $k$ ,  $\beta_{21} + b_{1k}^{(3)}$ . However, its presence in the time-to-event model through the association structure adjusts the population effect of treatment group assignment  $\beta_{21}$  to give a study specific estimate of deviation in log hazard ratio due to treatment assignment  $\beta_{21} + \alpha^{(3)} b_{1k}^{(3)}$ .

An advantage of this model is that as the number of studies included in the meta-analysis increases, the model does not become unwieldy, as the number of parameters that control the distribution of the study level random effects remains constant. However, the distribution of the study level random effects is based effectively on a sample size equal to the number of studies in the meta-analysis. The estimation of covariance matrix  $\mathbf{A}$  could be based on little



information if few studies contribute to the meta-analysis. As such, unless over a given number of studies is included in the meta-analysis, it may not be sensible to consider the estimated distribution of the study level random effects to be reliable. This could be an issue in the joint modelling case, as the study level random effect do not just account for between study heterogeneity, they act as part of the function linking the longitudinal and the time-to-event sub-models. Additionally, many meta-analyses contain less than 10 studies, which may not be sufficient to estimate these study level random effects. It is important to establish how reliably the distribution of the study level random effects is estimated during model fitting, an issue investigated further through simulations in Chapter 7.

Another consideration with this modelling group, as with group 2, is that study specific treatment coefficients are not directly estimated. However, when meta-analysing the data, the overall aim is to estimate a pooled treatment effect estimate based on data from all studies. This is accomplished in this model group through the estimation of  $\beta_{12}$  in the longitudinal sub-model and  $\beta_{21}$  in the time-to-event sub-model. The lack of study specific covariate estimates is generally not an issue in a meta-analytic investigation, and if they are required, a model group that uses fixed interaction terms should be employed. In addition, this model still assigns a common baseline hazard to all studies, thus assuming proportional hazards across all studies (an assumption identified to be potentially unreasonable).

### 3.3.2.5 *Group 4: Unspecified baseline hazard stratified by study, fixed interaction term with study membership variable in longitudinal sub-model*

The fourth group of models displayed in Table 5 introduces stratification of baseline hazard by study. The longitudinal sub-model of this group contains population fixed intercept, time ( $t_{kij}$ ), treatment assignment ( $treat_{ki}$ ), study membership ( $study_{ki}$ ) and interaction between treatment assignment and study membership terms (with coefficients  $\beta_{10}$  to  $\beta_{14}$ ), as well as individual level random intercept ( $b_{0ki}^{(2)}$ ) and random time ( $b_{1ki}^{(2)}$ ) terms and an error term ( $\varepsilon_{kij}$ ).

As before, the individual level random effects follow distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ , where  $\mathbf{D}$  is a 2 by 2 covariance matrix with on-diagonals equal to the variance of each random effect, and off-diagonal equal to their covariance. The random effects and the error terms are independently distributed. The error terms at each time point follow  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ .

The time-to-event sub-model contains a population fixed treatment assignment ( $treat_{ki}$ ) term with coefficient  $\beta_{21}$ , as well as an unspecified baseline stratified by study  $\lambda_{0k}$ . The association structure (fully stated in Table 5), involves just the individual level random effects.

As such, between study heterogeneity is accounted for in the longitudinal sub-model through the fixed study membership variable and the fixed interaction term between study membership and treatment assignment. Study specific estimates of the treatment effect can be generated through the sum of the treatment assignment coefficient for the baseline study ( $\beta_{12}$ ) and the study specific  $\beta_{14}$  variables resulting from the interaction between study membership and treatment effect. This study specific  $\beta_{14}$  coefficient quantifies the difference between the treatment effect in the baseline study and a given study in the meta-analysis for the longitudinal outcome. In this model, the presence of between study heterogeneity attributable to differences in treatment effect between studies can be tested for by comparing the differences in deviances between models containing and not containing interaction between study membership and treatment assignment. The difference in deviances will follow a  $\chi^2$  distribution with  $P$  degrees of freedom, where  $P$  is the difference in number of parameters between the two models.

Study specific longitudinal intercepts are obtained through the sum of the baseline study intercept  $\beta_{10}$  and the estimated study specific  $\beta_{13}$  variable linked to the dummy variable for the appropriate study. The study specific  $\beta_{13}$  coefficient quantifies the difference in the intercept of the longitudinal trajectory between the baseline study, and a given study in the meta-analysis. The residual between study heterogeneity in the longitudinal outcome not accounted for through the treatment assignment variable can be tested for by comparing the deviance between a model with and without study specific fixed intercepts.

In the time-to-event sub-model, between study heterogeneity is captured by the unspecified baseline, which is stratified by study. As such, between study heterogeneity is captured, but it is not estimated. As the aim of the meta-analysis is to provide a pooled coefficient estimate based on data from all studies, this may not be an issue unless it is necessary to explain existing heterogeneity. In this case, further covariates etc. could be included in the model.

Unlike previous models, by stratifying baseline hazard by study, proportional hazards are assumed within studies rather across all studies included in the meta-analysis. This may be a

more reasonable assumption than a common baseline hazard across studies especially if the demographics of the studies differ considerably.

Using study specific baseline hazards rather than a common baseline hazard may also reduce computation time, as the baseline hazard will take weight only at event times observed in each study rather than those observed in all studies, and so the length of the vector for baseline hazard will be shorter (and study specific). However, if use of a stratified baseline hazard does not fulfil the needs of the investigation, for example if the variability of a coefficient between studies is of interest, a stratified baseline hazard should not be employed just on the basis of computation time. Additionally, as number of studies increases, the computational burden of calculating a baseline hazard for each study may become relevant.

### 3.3.2.6 Group 5: Unspecified baseline hazard stratified by study, study membership variable in longitudinal sub-model, study level random treatment effect.

The fifth group of models described in Table 5 include a variety of methods to account for between study heterogeneity. The longitudinal sub-model contains a fixed intercept, longitudinal time term ( $t_{kij}$ ), treatment assignment ( $treat_{ki}$ ) and study membership variable ( $study_{ki}$ ), as well as two sets of random effects ( $\mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}$ ), and an error term ( $\varepsilon_{kij}$ ).

The individual level (level 2) random effects  $\mathbf{b}_{ki}^{(2)}$  contain a random intercept  $b_{0ki}^{(2)}$  and a random time term  $b_{1ki}^{(2)}$ , which again follow distribution  $\mathbf{b}_{ki}^{(2)} \sim N(\mathbf{0}, \mathbf{D})$ . The study term with coefficient (level 3) random effects  $\mathbf{b}_k^{(3)}$  only involves a random treatment effect  $b_{1k}^{(3)}$  which has distribution  $\mathbf{b}_k^{(3)} \sim N(0, \sigma_A^2)$ , where  $\sigma_A^2$  is a single value giving the variance of the study level random effect. The random effects at each level are considered independent, and are also independent of the error terms. The error terms are identically and independently distributed at each time point with  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ .

The time-to-event sub-model contains a fixed treatment assignment with coefficient  $\beta_{21}$ , as well as a baseline hazard stratified by study. The association structure (shown in Table 5) involves both individual and study level random effects, with common association parameter  $\alpha^{(2)}$  for the shared individual level random effects, and common association parameter  $\alpha^{(3)}$  for the study level random effects. As before, longitudinal time in the association structure is replaced with individuals' survival time.

In this group of models, between study heterogeneity is accounted for in a range of ways. In the longitudinal sub-model, the fixed study membership variable generates fixed study specific longitudinal intercepts, through the sum of the intercept for the baseline study  $\beta_{10}$  and the study specific coefficient for the relevant study's dummy membership variable  $\beta_{13k}$ .

Additionally between study heterogeneity due to treatment assignment is quantified through the study level random treatment effect  $b_{1k}^{(3)}$ . The level of between study heterogeneity in the treatment effect can be assessed by examining the variance of the random effect  $\sigma_A^2$  (which in this case is a single value) in the same way as for model group 2 (comparing models with and without study level random effects: if the model with study level random effects displays lower information criteria values then there is evidence of between study heterogeneity in treatment effect). The residual between study heterogeneity not accounted for by variation in treatment effect can be assessed by the difference in the deviance between joint models with and without a fixed study membership variable in the longitudinal sub-model.

Between study heterogeneity is accounted for in two ways in the time-to-event sub-model. Firstly, as for group 4, the baseline hazard is stratified by study. As such, by allowing the baseline hazard to vary between studies, the between study heterogeneity is accounted for but not quantified. Through the association structure, the term  $\alpha^{(3)}b_{1k}^{(3)}$  occurs in the time-to-event model. This term adjusts the log ratio hazard of risk of an event dependent on treatment assignment from the population estimate  $\beta_{21}$  by an amount specific to each included study.

The magnitude of the adjustment due to study level random effects can be examined in the same way as for model group 2. Specifically, once the model has converged, potential study specific treatment effect parameter estimates can be generated from normal distributions based on the estimated model parameters. Let the estimated distribution of the study level random treatment effect  $b_{1k}^{(3)}$  be  $N(0, \hat{\sigma}_A^2)$ . To assess possible study specific longitudinal treatment effect estimates under this model, generate realisations from a  $N(\hat{\beta}_{12}, \hat{\sigma}_A^2)$ . To assess possible study specific time-to-event treatment effect, generate realisations from a  $N(\hat{\beta}_{12}, (\hat{\alpha}^{(3)})^2 \hat{\sigma}_A^2)$ . Histograms of the realisations can help to visualise the mean, range and spread of potential study specific estimates from hypothetical generated studies; the more spread out the values, the greater the between study heterogeneity.

Between study heterogeneity in this model may be more difficult to interpret than the other models, because it is accounted for in a variety of ways. In addition, as number of studies increases, this model may experience issues due to the necessity to calculate a separate baseline hazard for each included study, and also the separate coefficients for study membership estimated for all but the baseline study. However, as with group 4, model fitting times are likely to be improved by the smaller vector lengths of the baseline hazard, as it will take weight only at event times observed in each study rather than those observed in all studies.

### 3.3.3 Number of parameters in one-stage joint MA models

In the descriptions of model groups 0 through 5, it was mentioned when fixed effects are used to account for between study heterogeneity, presence of between study heterogeneity can be tested by comparing the differences in deviances between models containing and not containing the relevant fixed effects (for example comparing the deviances of a model containing fixed interaction terms between study membership and treatment assignment, and a model without these interaction terms, to test for between study heterogeneity in treatment effect). The difference in deviances has been noted to follow a  $\chi^2$  distribution with  $P$  degrees of freedom, where  $P$  is the difference in number of parameters between the two models. To calculate  $P$ , the number of parameters estimated in each model must be obtained. This is easily calculated as follows:

Classing baseline hazard as a nuisance parameter (and so not counting it towards parameters to be estimated), let  $P_M$  be the number of parameters in a given model  $M$ .  $P_M$  can be expressed:

$$P_M = p_1 + p_2 + \varrho + \frac{1}{2}q(q + 1) + \frac{1}{2}r(r + 1) + 1$$

In the above expression,  $p_1$  is the number of fixed effects in the longitudinal sub-model, and  $p_2$  is the number of fixed effects in the time-to-event sub-model. In each sub-model, the number of fixed effects includes any interaction terms present, and assumes factors have been decomposed into dummy variables. The number of association parameters to be estimated is given by  $\varrho$ . Under the joint models examined in this chapter, if the model only contains individual level random effects,  $\varrho = 1$ , and if the model contains both individual level and study level random effects,  $\varrho = 2$ . The number of individual level random effects is denoted

by  $q$ , and the number of study level random effects is denoted by  $r$ . The number of parameters to be estimated for each set of random effects is the number of unique values in the symmetric covariance matrix for the relevant random effect distribution. This is equal to the total of the number of on-diagonal parameters and the number of unique off-diagonal parameters, i.e. the  $n$ th triangular number where  $n$  is equal to the number of random effects at the given level. Note, if no random effects are included at a given level (e.g. if no study level random effects are present in the model meaning  $r = 0$ ) then the corresponding term in the expression for  $P_M$  also equals zero (i.e.  $\frac{1}{2}r(r + 1) = 0$ ). Finally, the plus one at the end of the expression represents the parameter quantifying the variance of the measurement error (namely  $\sigma_e^2$ ). The difference in number of parameters between two models  $M_1$  and  $M_2$  can then be calculated as:

$$P = P_{M_1} - P_{M_2}$$

Here, model  $M_1$  contains the larger number of parameters.  $P$  then gives the degrees of freedom for the  $\chi^2$  distribution, which the differences in deviance between model  $M_1$  and model  $M_2$  should be compared to, in order to assess the presence of between study heterogeneity as accounted for using the fixed terms present in model  $M_1$  but not model  $M_2$ .

In the future, if alternative formulations for the baseline hazard are used (e.g. parametric or spline based) the number of parameters required for their estimation should contribute to the calculation of  $P$  (see Chapter 8 for planned future work).

### 3.4 Discussion of meta-analytic joint modelling methods

During this chapter a range of methods for the one-stage or the two-stage meta-analysis of multi-study joint longitudinal and time-to-event data have been presented and discussed. It was assumed that IPD is available to the researcher, with a note that the second stage of the two-stage methods could be used if performing an AD-MA of joint data.

#### 3.4.1 Discussion of methods for two-stage joint IPD-MA

During Section 3.2, the procedure of fitting joint models to each identified study, extracting the parameters and quantities of interest, and pooling the results using standard meta-analytic techniques has been described. In addition, guidelines for the conduct of the meta-analysis were presented, highlighting procedures to ensure that parameters with differing

interpretations are not quantitatively pooled, thus ensuring that pooled parameters have meaningful interpretations.

A two-stage MA of joint models was possible at the start of this thesis using currently available software. In R [79], for the first stage, various packages exist to fit frequentist single study joint models, including the *joiner* package [80], and the *JM* package [81]. Whilst not directly designed to model single study univariate models, the *joinerML* package [181], and the *joinerMeta* package (which I have written during this thesis, available on github at <https://github.com/mesudell/joinerMeta>, and for download from the R CRAN mirror), both can also fit single study univariate joint models. In the second stage a range of packages exists to perform meta-analyses, including the *meta* [182] and *metafor* [183] packages. Software also exists in other packages (such as Stata [84] or SAS software [85]). However, the process to extract the values necessary for a meta-analysis from a joint model fit in R, and feed them into a meta-analysis package is convoluted, especially for those with limited programming experience. To address this, in the *joinerMeta* package, I have included a function that, when supplied with joint model fits from *joiner* or *JM*, along with names of parameters of interest, extracts and performs the second stage of the two-stage MA process (see Chapter 4).

During the investigation of methods for the two-stage MA of joint data, it was assumed that separate MA would be performed for each parameter of interest. However, this could be a naïve approach. In a joint model, a core concept is that the longitudinal and time-to-event sub-models are linked and so affect each other. As such, given a significant estimated association between the longitudinal and the time-to-event outcome, there would be correlation between parameters from various parts of the joint model. Performing separate MA for each parameter of interest ignores this correlation. A solution to this could be to perform multivariate meta-analyses. This extension is further discussed in Chapter 8.

The difference in interpretation of different association structures in joint models available to researchers leads to an important concept to consider in both two-stage IPD-MA, and in aggregate data meta-analyses (AD-MA, where study level results are obtained from published information or other sources, and pooled). During a meta-analysis involving joint models this research recommends that the association structure should be kept consistent across included studies, in order to be able to pool the association estimates. If the most appropriate joint modelling structure (random effect specification, association structure, longitudinal fixed

effects if both fixed and random effects are shared between sub-models) differs between identified studies, studies should be grouped by joint model structure, and only results from identical joint models (within group results) should be pooled. Pooled estimates from different groups (different joint modelling structures) can then be qualitatively compared in the discussion of the meta-analysis, whilst bearing in mind their potentially different interpretation. Recommendations for the two-stage MA of joint longitudinal and time-to-event data have been succinctly stated in Section 3.2.4.

The recommendations for two-stage MA of joint longitudinal and time-to-event IPD (Section 3.2.4, second stage applicable to AD) do not state that results should not be pooled from joint models employing different methods to model baseline hazard. The recommendations have been produced to aid researchers in avoiding pooling parameters with non-comparable interpretations. Generally in meta-analyses of time-to-event data, coefficients quantifying the difference in risk between patients with different characteristics are of interest, for instance treatment assignment coefficients. A proportional hazards time-to-event sub-model (for example) that uses an unspecified baseline hazard will produce fixed effect coefficients with comparable interpretations to one that employs a spline based or parametric baseline hazard. The method chosen to model the baseline hazard should be able to account for any characteristics of the baseline hazard. For instance, if the baseline hazard initially increases, plateaus, then decreases, the method used to represent the baseline hazard must be able to represent this shape; modelling the baseline hazard using a constant function in this case would be inappropriate. However, unless the baseline hazard itself is of interest to the analysis, provided it is appropriately modelled, and the fixed effect coefficients themselves have comparable interpretations, the coefficients can be pooled.

The review detailed in Chapter 2 identified that joint modelling analyses employed a mix of frequentist and Bayesian approaches (Section 2.3.1.7). It should be highlighted that the modelling assumptions of Bayesian and frequentist approaches are inherently different (see Section 16.8.1, Cochrane Handbook [184]). As such, it is not appropriate to pool parameters from a mix of Bayesian and frequentist analyses. The choice of Bayesian or Frequentist modelling approach should be made before the first stage of analysis for a two-stage IPD-MA. For an AD-MA, Bayesian and Frequentist analysis results should be pooled separately. The



conclusions based on these two sets of analyses can then be qualitatively compared and discussed.

The methods discussed in this chapter assume availability of IPD. It is possible that in practice, IPD could be obtained only for a subset of studies eligible for inclusion in the meta-analysis, with AD available for the remaining references. The review conducted by Riley et al [3] identifies three approaches to pooling IPD and AD, namely using multi-level modelling, using Bayesian hierarchical related regression, and using a two-stage approach similar to that described in this chapter. Under this approach, in the first stage, study level estimates would be obtained from any IPD data, and in the second stage these estimates would be combined, along with any AD extracted, to generate a pooled result. In the context of a joint data MA, care would need to be taken to ensure that the joint modelling methods used to reduce the IPD to study specific AD were comparable to the methods stated as used to produce the AD extracted from studies, in line with the recommendations for conduct of two-stage MA of joint data (see Section 3.2.4).

It is possible that the data supplied by each study in a meta-analysis could contain different subsets of potential treatment effect modifiers. In this circumstance, for a two-stage joint IPD-MA, covariates should be selected separately for each study, using forward or backward selection methods, and an estimate of treatment effect after adjustment for covariates that significantly affect the longitudinal or the time-to-event outcomes should be output for each study. Sensitivity analyses can then be performed, removing studies that did not adjust for certain covariates, to assess the robustness of the pooled result. Information on the covariates adjusted for in each study could be stated on forest plots summarising the conducted meta-analyses.

As an aside, in the case of AD-MA, where the results are comparable, study level results could be pooled, and the reported adjustment covariates noted e.g. on the forest plot displaying the pooled data. Again, sensitivity analyses could be performed for AD-MA, removing studies that do not adjust for potential treatment effect modifiers, or removing studies that only adjust for a subset of the potential treatment effect modifiers, to test the robustness of the conclusion.

### 3.4.2 Discussion of methods for one-stage joint IPD-MA

During Section 3.3, a range of model groups to model multi-study joint longitudinal and time-to-event data were presented. Each of these model groups accounted for between study heterogeneity in different ways, through varying combinations of fixed terms (including study membership and its interaction with covariates of interest), study level random effects, and stratification of the baseline hazard in the time-to-event sub-model. The different methods to account for between study heterogeneity between model groups have been discussed, along with their advantages and drawbacks. Functions to fit the models described in this section have been included in the *joinerMeta* package, discussed in Chapter 4.

As for the two-stage joint data MA (Section 3.2.1), in a one-stage joint MA it would be advisable before conducting the analysis to produce plots of the longitudinal outcome trajectories panelled by event. These trajectory plots could also be panelled by other covariates of interest, e.g. treatment group, to provide an initial representation of whether the behaviour of the longitudinal outcome differs between different groups. Plotting the longitudinal outcome against the longitudinal time variable can help to identify aspects of the trajectory shape that need to be accounted for (e.g. non-linear behaviour). Plotting the longitudinal outcome against time adjusted by survival time can identify whether the longitudinal outcome changes immediately before experiencing an event— this would provide evidence that the longitudinal and time-to-event outcomes may be linked, implying joint modelling techniques are required. Additionally, producing Kaplan-Meier plots for the time-to-event outcome is advised, e.g. so that assumptions of proportional hazards within or across studies could be checked. The longitudinal trajectory and the Kaplan-Meier plots would inform the choice of model specification for the longitudinal sub-model, time-to-event sub-model and the association structure.

One issue identified is the suitability of different methods dependent on the number of studies included in the meta-analysis. As stated, if a large number of studies is available, a model using fixed terms such as the study membership variable and its interactions with other covariates will contain a large number of parameters to be estimated, potentially leading to an unwieldy model. However, if only a few studies are available, the distribution of any study level random effects could be poorly estimated. Consequently, effect of the number of studies on estimation

of model parameters in different model groups is investigated during the one-stage simulation study (Chapter 7).

In addition to considerations about the actual estimation of the model parameters, models that employ fixed study membership and its interaction with covariates of interest terms provide parameter estimates specific to the studies included in the meta-analysis. As such, generalisation of the results to a wider population outside the included studies could be problematic. Alternatively, model group 3 relies completely on study level random effects to account for between study heterogeneity. The distribution of study level random effects, which quantifies the level of between study heterogeneity, is estimated during model fitting. As long as the studies included in the meta-analysis are representative of the wider population, results can be easily generalised given this distribution. However, study level random effects, whilst accounting for study heterogeneity, do not automatically provide estimates for study specific effects; this is not generally a main aim of meta-analyses and so is not a major drawback.

Whilst joint modelling has been an area of increasing research [74-76, 93], the development of model diagnostics or procedures to choose the most appropriate model has been somewhat slower. Recently, publications have started to appear dealing with this area, including Rizopoulos et al [185] who discuss using residuals as a joint model diagnostic tool. However, their methods exclude joint models that involve a time-to-event sub-model with an unspecified baseline hazard. In addition, Park et al [186] discuss model selection and diagnostics for joint models, but again restrict themselves to the case where there are crossing hazard functions. Dobson and Henderson [187] present a range of graphical methods to assess joint models, however note that they are informal assessments rather than formal tests.

Rizopoulos [52] describes a range of methods to test the inclusion of various parameters in a joint model, including the Likelihood Ratio Test, the score test and the Wald test. However, these tests are appropriate only when the models are nested (and as discussed earlier can be used to test for between study heterogeneity when accounted for using fixed effects).

As stated, to compare non-nested models (e.g. those from different groups), information criteria such as Akaike's Information Criterion (AIC) [178], or the Bayesian Information Criterion (BIC) [179] could be used [52, 77]. For these values, the smaller values of the AIC/BIC identify the better model. Calculation of the AIC value has been programmed into the output from the

one-stage modelling function included in the *joinermeta* package, to facilitate comparison of different one-stage joint MA models. This is further discussed in Chapter 4.

If potential treatment effect modifiers are supplied in the dataset, these should be examined in the model, and retained where significant. Covariates can be assessed through forward or backward selection processes. Ideally, the same treatment effect modifiers would be measured in all studies. This may not always be the case, and so inclusion of certain treatment effect modifiers may remove some individuals, or even entire studies, from the meta-analyses. A range of methods to account for this exist, such as performing sensitivity analyses where multiple imputation techniques are used to impute missing covariates for individual or studies, based on the observed data, and comparing the analyses based on imputed data to those based on the recorded dataset. There has been a recent movement towards developing core outcome sets for different conditions, led by the COMET initiative [188]. This initiative aims to produce sets of outcomes that should be measured as standard in each study. This may result in lower amounts of missing data for covariates of interest for individuals and studies in future meta-analyses.

### 3.4.3 Concluding Remarks

This chapter has presented and discussed methodology to conduct both one and two-stage meta-analyses. These two approaches have differing characteristics. One-stage approaches employ more accurate likelihood specifications than two-stage approaches [189-191], but care must be taken to separate within and between study effects [180]. Two-stage methods automatically account for the issue of ecological bias as they only pool within study information [63, 180, 192], but they assume that the study specific estimates are normally distributed with known variances (a potential issue for studies with small sample sizes or applications to rare events) [191, 193]. One-stage methods are known to be computationally intensive [63, 189, 191]; one-stage analyses involve fitting models to  $\sum_{k=1}^K n_k$  individuals, whereas two-stage analyses involve fitting  $K$  models, each of which involves  $n_k$  individuals. The results of a one and two-stage meta-analysis are often similar, but differences between results can occur for a variety of reasons [189]. Debray et al [194] and Burke et al [189] recommend that if researchers are unsure which approach to employ, both a one and a two-stage analyses should be planned, conducted and reported.

This chapter has proposed methods to model multi-study joint data in meta-analyses. In practice, there may be many covariates that modify the treatment effect parameter, which should be included in the model. Investigation of patient level treatment effect modifiers is only possible when IPD is available, without results being subject to ecological bias [9], and so may be a motivation for AD rather than IPD-MA if treatment modifiers are known or suspected. In two-stage IPD analyses, modifiers can be included in the study specific joint models, however they will only represent the effect of the covariate on treatment effect based on the range of covariate values reported in that study. If demographics of included studies differ significantly between studies, sub-group analyses may be required in the second stage that group studies into specified categories (e.g. age  $\leq 50$ , age  $>50$ ) and pool results only within these sub-groups. In one-stage analyses, data from all comparable studies can be analysed under the same model, with differences in demographics accounted for by inclusions of relevant covariates in the model. Accounting for differences in the demographics between included studies is established across meta-analysis methodology, not limited to joint data meta-analyses. Analyses conducted with an aim to influencing future healthcare should firstly assess the demographics of the identified studies, and if the populations are not too disparate, should employ modelling approaches that account for differences in population demographics.

In the literature, meta-analyses may be expected to be updated over the years, as new studies relevant to the meta-analysis become available. In this case, the meta-analysis would have to be updated with respect to the new data (a process observable in Cochrane Reviews). If a two stage approach to the MA had been undertaken, and the study level results for the currently included studies preserved, the analysis could be quickly updated by re-performing the second or pooling stage of the meta-analysis. However, the process of updating a one-stage analysis would be more involved, as for the frequentist approach described in this chapter, the entire analysis would have to be redone. An alternative to this (not investigated in this thesis) would be a Bayesian approach, where the information from currently included studies would be preserved in the prior of the new analysis, and the newly available data then included in the likelihood. However, as mentioned in the Cochrane Handbook [184] (section 16.8.1), methodology for Bayesian meta-analysis is still being developed and is not yet widely implemented.

Having presented methods for the meta-analysis of multi-study joint data, Chapter 4 presents and discusses the software developed during this thesis to aid the meta-analysis of joint data. Over the chapters that follow, an analysis of a real dataset is presented, examining the feasibility of the methods proposed in this chapter in a real dataset (Chapter 6). This is followed by an in depth analysis of the behaviour of the proposed methods (both one and two-stage) under a range of different scenarios such as varying levels of association and heterogeneity (Chapter 7).

## Chapter 4 : R Software package for Meta-Analysis of Joint Longitudinal and Time-To-Event Data

---

This chapter presents and discusses the R package *joineRmeta* that I have developed to facilitate researchers when conducting meta-analyses of joint longitudinal and time-to-event data. The current developmental version is available at <https://github.com/mesudell/joineRmeta/> (where updates to the package will be loaded and tested before being included in the CRAN version of the package). The current tested version of the package is available for download from CRAN in R [79] directly through R software interfaces. I am listed as the maintainer of the package. The chapter begins with a discussion of the simulated data and pre-run model fits available in the package for demonstration purposes, and is followed by details of the various functions contained in the package. This discussion is split into Analysis Functions (including the functions for one-stage or two-stage MA of joint data), followed by Exploratory Functions (including data simulation, preparation and visualisation). The chapter concludes with a discussion of the R package, and its use in the remainder of the thesis.

During the chapter, reference will be made to R functions or code which can be run in the console in R. Code is clearly marked during text by a change in font e.g. *jointmeta1* or is displayed in blocks between paragraphs, e.g.

```
jointmeta1()
```

Throughout, example code is used to demonstrate the functionality of the package, and the output is presented as it would appear when code is run in the R console. As such, graphs in this chapter are presented as they would appear in the output and have not been assigned captions.

In the text, when a function is introduced, the function call is printed. The default values for arguments, or the possible values the argument can take (if these are limited), are displayed in the function call or discussed in the main text. After the function call, brief definitions of the arguments involved in the function are given, with further discussion of function options after the argument definitions where appropriate.

It is expected that names of data variables are supplied as character strings. For example, if the name of an individual identification or ID variable named *IDVAR* were being supplied to argument *id* in a function, the phrase *id = "IDVAR"* should be included in the function call. Using phrase *id = IDVAR* would cause R to search for an object named *IDVAR* to supply to the function.

A range of models have been fitted to example data to demonstrate the functions in this package. These models may not necessarily represent the most appropriate model for the data.

## 4.1 Installing the *joineRmeta* package

The developmental version of the *joineRmeta* package can be loaded in R using the following commands:

```
library(devtools)
install_github("mesudell/joineRmeta")
library(joineRmeta)
```

The fully tested version of the package, available through the CRAN mirror, can be installed using:

```
install_packages("joineRmeta")
library(joineRmeta)
```

Information concerning the functions and their usage is available in the help files and the user-friendly vignette (tutorial file) of the *joineRmeta* package (available at the mentioned web address and by downloading the R package). When the package is loaded in R, the help file for any function can be accessed by typing *?functionname* into the R console, e.g.

*?jointmeta1*. Additionally, the help files overall for the package can be accessed using:

```
help(package="joineRmeta")
```

## 4.2 Data

For demonstration purposes, the *joineRmeta* package contains simulated datasets and pre-run joint model fits to allow the user to explore the package's functionality without having to wait for lengthy model fitting processes such as bootstrapping to complete. These datasets and



fits are used in the examples throughout this chapter, but play no part in other chapters. A brief description of the examples used follows.

#### 4.2.1 Simulated data

The *joinerMeta* package contains three simulated example datasets, named *simdat*, *simdat2* and *simdat3*, generated using the *simjointmeta* function contained in the package (see Section 4.4.1). These datasets contain a single continuous longitudinal outcome and a single time-to-event outcome. Datasets *simdat* and *simdat3* contain 5 studies, and *simdat2* contains data from 3 studies. The datasets are supplied as a list of 3 list objects; a list of study specific longitudinal datasets *Longitudinal*, a list of study specific time-to-event datasets *survival*, and a list of the event rates in each study *percentevent*.

The *simdat2* dataset is a subset of *simdat* included to provide datasets containing different numbers of studies for use with example code. Both the *simdat* and *simdat2* datasets were generated such that longitudinal measurements made after the individuals survival time are not present in the dataset, however *simdat3* contains longitudinal measurements recorded after an individual's event time. Real datasets may contain longitudinal measurements recorded after the event if the event is not terminal, and so a function has been included in the *joinerMeta* package to remove longitudinal data recorded after an individual's event time (see Section 4.4.2.2). The *simdat3* dataset is included in the package in order to demonstrate this function.

Further information concerning these datasets is available by loading the *joinerMeta* package into R, and typing *?simdat*, *?simdat2* or *?simdat3* into the console.

#### 4.2.2 Single study joint model fits

The function *jointmeta2* is written to aid with the second stage of two-stage MA of joint data. This function is described fully in Section 4.3.1. The function can take a list of study specific joint model fits, and meta-analyse model parameters. The function can take fits from either the *joiner* or the *JM* packages. Example fits from each package are supplied in objects *joinerRfits*, *joinerRfits2*, *JMfits* and *JMfits2*. As with the simulated datasets, more information can be obtained by loading the *joinerMeta* package in R, and typing e.g. *?joinerRfits* into the console. The model fits represent differing options for random effects

or fixed effects specifications, as well as different choices of association structure. As the model fits produced by *joiner* contain both the model fits, and the results of the bootstrapping procedure to obtain standard errors, these need to be extracted before conducting the meta-analysis, using the following code:

```
joinerModels <- joinerFits[c("joinerFit1", "joinerFit2", "joinerFit3")]
joinerModelsSE <- joinerFits[c("joinerFit1SE", "joinerFit2SE", "joinerFit3SE")]

joinerModels2 <- joinerFits2[c("joinerFit6", "joinerFit7", "joinerFit8")]
joinerModels2SE <- joinerFits2[c("joinerFit6SE", "joinerFit7SE", "joinerFit8SE")]
```

#### 4.2.3 One-stage fits to multi-study joint data

The function *jointmeta1* is written to allow the one-stage analysis of multi-study joint data. As this function requires bootstrapping to obtain standard errors, example model fits and bootstrapping results are available in the package, in objects *onestage0*, *onestage1*, *onestage2*, *onestage3*, and *onestage4*. The provided one-stage fits display a range of the modelling options available, including study level random effects, baseline hazard stratified by study, or fixed study membership terms and interactions between study membership and covariates of interest. Again, as these objects are not directly used in the thesis, but are present in the package to aid future users to familiarise themselves with the available functions, they are not discussed any further here. Additional information can be obtained by loading the *joinermeta* package in R, and typing, e.g. `?onestage0` into the console.

### 4.3 Analysis Functions

The functions discussed in this section relate to the meta-analysis of multi-study joint data. The section is split into two sections, the first presenting functions to implement two-stage MA of joint data, and the second describing functions to implement one-stage MA of joint data. For details of proposed methods to conduct one or two-stage MA of joint data, see Chapter 3.

#### 4.3.1 Methods for Two-stage Meta-Analyses

The function *jointmeta2* can take the results of joint model fits from the *joiner* [80] and the *JM* [81] packages, extract the specified parameters of interest and perform separate meta-analyses for each parameter. This function simplifies the process of a two-stage MA for joint

longitudinal and time-to-event data by reducing the level of programming required to perform the meta-analysis. The *jointmeta2* function has the following syntax:

```
jointmeta2(fits, SE = NULL, longpar = NULL, survpar = NULL,  
  assoc = TRUE, studynames = NULL)
```

The arguments of the function have the following definitions:

- ***fits*** - a list of joint modelling fits. These fits should all be of the same type (i.e. all fitted using *joint* function from the *joiner* package or all fitted using the *jointModel* function from the *JM* package), with the same model specification.
- ***SE*** - a list of the results from the *jointSE* function from the *joiner* package (only to be supplied if the model fits supplied to argument *fits* are all fitted using the *joiner* package)
- ***Longpar*** - a vector of names of parameters from the longitudinal sub-model for which meta-analyses should be performed
- ***survpar*** - a vector of names of parameters from the survival sub-model for which meta-analyses should be performed
- ***assoc*** - a *TRUE/FALSE* indicating whether a meta-analysis should be performed for the association parameter(s)
- ***studynames*** - a vector of the names of the studies present in the dataset that the joint models were fitted to. These are used to label the meta-analyses performed by the function

This function will perform both a fixed and a random effects MA. Meta-analyses are performed using the *meta* package [182] in R, and so the results of the function can be fed into the *forest* function in the *meta* package to produce forest plots of the analyses.

The *jointmeta2* function will return a list containing lists of meta-analyses, whose elements depend on the package used to produce the original joint model fits. This is due to the differing association structures employed by the *joiner* and the *JM* packages. The *joiner* package employs random effects only association structures, i.e. zero mean random effects are shared between the longitudinal and time-to-event sub-models to link them. In comparison, the *JM* package employs association structures that rely on some function of both the fixed and the random effects included in the longitudinal sub-model to link the sub-models. As discussed in

Chapter 3, pooling of parameters from joint models of different specifications is not advised as the parameters have different interpretations. To prevent this from happening, the function does not allow results from a mix of JM and joineR fits to be pooled, and code has been included in the function to check the specifications of the joint model fits supplied.

If the models supplied to the *jointmeta2* function all had the same specification, the function would proceed with the meta-analysis. In the example below, the supplied joint models all employ a joint model whose longitudinal sub-model contains a fixed intercept, time and treatment terms and random intercept and time terms, and whose time-to-event sub-model contains a fixed treatment effect. The sub-models are linked through shared random effects with common association parameter. The function call (printed below) specifies that meta-analyses are performed for the time and treatment terms in the longitudinal sub-model (using argument *Longpar* = *c("time", "treat1")*), for the treatment effect in the time-to-event sub-model (with argument *survpar* = *"treat1"*), and for the association parameter (through *assoc* = *TRUE*). Labels are also supplied to identify the studies contributing data to the analysis.

```
MAjoineRfits <- jointmeta2(fits = joineRmodels, SE = joineRmodelsSE,
  longpar = c("time", "treat1"),
  survpar = "treat1", assoc = TRUE,
  studynames = c("Study 1", "Study 2", "Study 3"))
```

The structure of the output is examined using the following code:

```
names(MAjoineRfits)
## [1] "longMA" "survMA.direct" "assocMA"
```

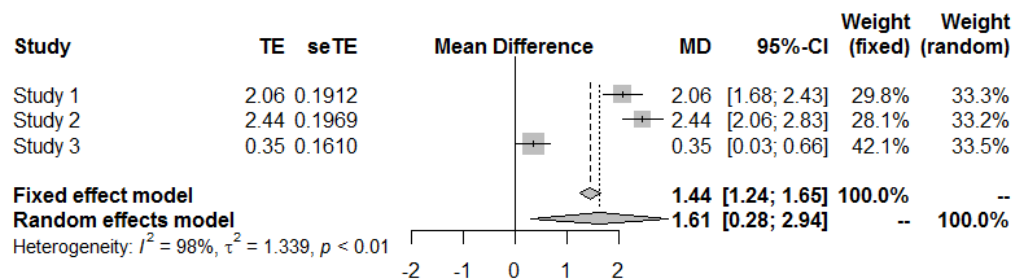
The output is a list containing three elements, each of which is a list of the results of meta-analyses performed using the *meta* packages for the parameters of interest for the longitudinal sub-model, the time-to-event sub-model and the association structure respectively. For example, the meta-analyses conducted for each of the longitudinal parameters of interest can be extracted using the code *MAjoineRfits\$LongMA*, or by using *MAjoineRfits\$LongMA\$treat1*, just the meta-analysis for the longitudinal treatment effect can be extracted:

```
MAjoineRfits$longMA$treat1
##           MD           95%-CI %W(fixed) %W(random)
## Study 1 2.0558 [1.6810; 2.4305]      29.8      33.3
```

```
## Study 2 2.4411 [2.0552; 2.8270]      28.1      33.2
## Study 3 0.3452 [0.0296; 0.6607]      42.1      33.5
##
## Number of studies combined: k = 3
##
##              MD          95%-CI      z  p-value
## Fixed effect model  1.4447 [1.2400; 1.6493] 13.84 < 0.0001
## Random effects model 1.6103 [0.2845; 2.9361]  2.38   0.0173
##
## Quantifying heterogeneity:
## tau^2 = 1.3390; H = 6.42 [4.63; 8.91]; I^2 = 97.6% [95.3%; 98.7%]
##
## Test of heterogeneity:
##      Q d.f.  p-value
## 82.46    2 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - DerSimonian-Laird estimator for tau^2
```

Forest plots can also be produced by:

```
library(meta)
forest(MAjoinerfits$longMA$treat1)
```



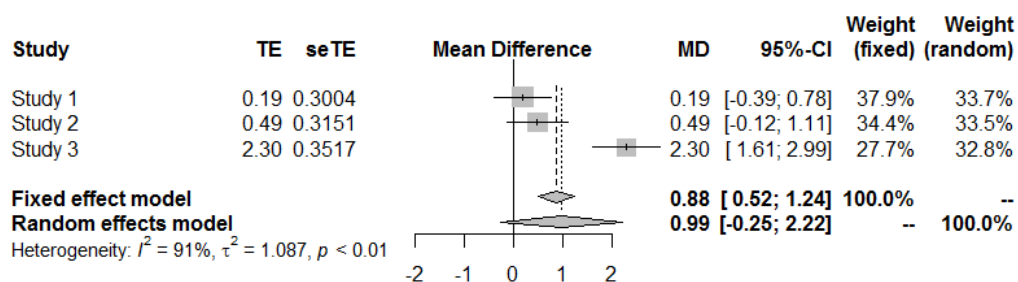
In the meta-analyses of fits from the *joiner* package, the meta-analyses of the time-to-event parameters are held in a list labelled *survMA.direct* while from the *JM* package these are held in *survMA.direct* and *survMA.overall*. This is due to the difference in available association structures between the two packages: the *JM* package allows association structures involving both fixed and random effects, whereas the *joiner* package shares only the random effects between the sub-models. As a result, in joint model fits from the *JM* package, covariates included in the time-to-event sub-model have an overall effect on risk of an event consisting of the direct effect of including the covariate in the time-to-event sub-model, and the indirect effect of fixed effects present due to the association structure. The theory behind this concept is described in Section 5.1.1.

Due to the presence of direct, indirect and overall effects in joint models that employ fixed and random effects association structures, when such joint modelling fits are supplied to the *jointmeta2* function, meta-analyses are performed for both the direct effect, and the overall effect of specified parameters from the time-to-event sub-model:

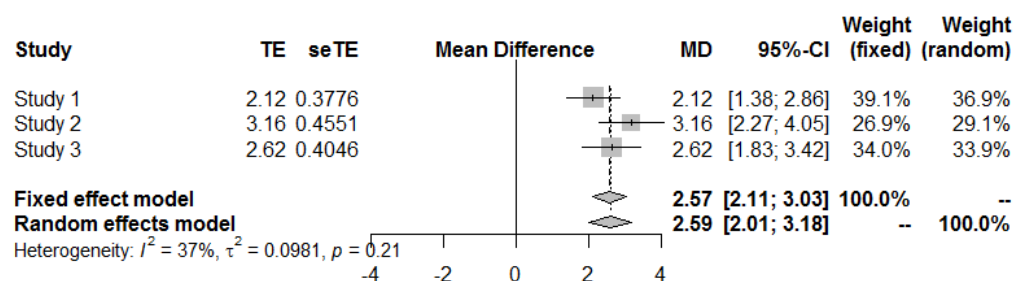
```
MAJMFits <- jointmeta2(fits = JMfits, longpar = c("time", "treat1"),
  survpar = "treat1", assoc = TRUE,
  studynames = c("Study 1", "Study 2", "Study 3"))
names(MAJMFits)
## [1] "longMA" "survMA.direct" "survMA.overall" "assocMA"
```

As before, forest plots of the results can be examined, allowing clear comparison of the direct and overall estimates:

```
forest(MAJMFits$survMA.direct$treat1)
```



```
forest(MAJMFits$survMA.overall$treat1)
```

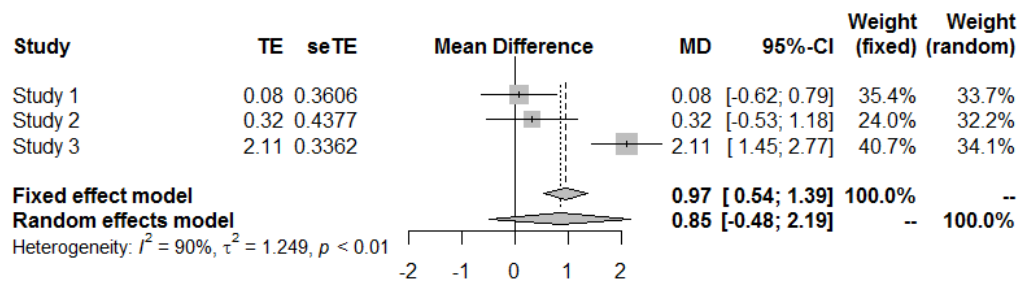


As well as being able to calculate the overall effects for models that use just a current value association structure, the function can also handle the case where the association structure of the joint model involves both the current value and current slope (first derivative with respect to time) of the longitudinal trajectory. The models included in object *JMfits2* use an association structure that involves both the current value and slope of the longitudinal

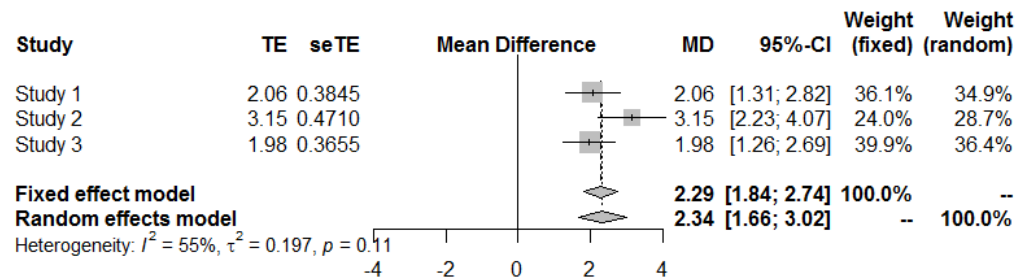
trajectory, and the longitudinal sub-model contains both a treatment assignment fixed effect and an interaction between time and treatment. As such the overall effect of the covariate treatment assignment on the risk of an event has format consisting of a direct effect resulting from inclusion of treatment as a fixed effect in the time-to-event sub-model, and one indirect effect from each of the current value and the current slope association structures (see Section 5.1.1 for in-depth discussion of the theory of indirect and direct effects in joint models). Meta-analyses can be performed by supplying the joint modelling fits to the *jointmeta2* function, and specifying the parameters of interest. The function will identify any coefficient estimates that should contribute to overall effects of specified time-to-event parameters of interest, and will calculate the overall value. This is demonstrated through the following code:

```
MAJMFits2 <- jointmeta2(fits = JMfits2, longpar = c("time", "treat1"),
  survpar = "treat1", assoc = TRUE,
  studynames = c("Study 1", "Study 2", "Study 3"))

forest(MAJMFits2$survMA.direct$treat1)
```



```
forest(MAJMFits2$survMA.overall$treat1)
```



As mentioned, based on the guidelines given for two-stage MA in Section 3.2.4, the function checks if the specifications of the joint models are identical. Various differences in supplied joint model specifications can cause error messages to be returned by the *jointmeta2* function.

Firstly, if the *jointmeta2* function is supplied with a mixture of models that link sub-models through shared random effects, and models that link sub-models through shared fixed and random effects, the following message is returned:

```
MAtest <- jointmeta2(fits = c(JMfits2[1:3], joinerModels[1:2]),
                     longpar = c("time", "treat1"),
                     survpar = "treat1", assoc = TRUE,
                     studynames = c("Study 1", "Study 2", "Study 3", "Study 4", "Study 5"))
## Error in jointmeta2(fits = c(JMfits2[1:3], joinerModels[1:2]), longpar = c
## ("time", : Some of the joint modelling fits are different classes -
##      consider subgrouping
```

Secondly, pooling model fits from the *joiner* package whose random effects specification differ will cause the following message:

```
MAtest <- jointmeta2(fits = c(joinerModels[1:3], joinerModels2[1:2]),
                     SE = c(joinerModelsSE[1:3], joinerModels2SE[1:2]),
                     longpar = c("time", "treat1"), survpar = "treat1",
                     assoc = TRUE,
                     studynames = c("Study 1", "Study 2", "Study 3", "Study 4", "Study 5"))
## Error in jointmeta2(fits = c(joinerModels[1:3], joinerModels2[1:2]), SE =
## c(joinerModelsSE[1:3], : Some of the joint model fits have differing
##      random effects structures
```

Additionally, pooling model fits from the *JM* package whose parameter specifications differ results in the following:

```
MAtest <- jointmeta2(fits = c(JMfits2[1:3], JMfits[1:2]),
                     longpar = c("time", "treat1"),
                     survpar = "treat1", assoc = TRUE,
                     studynames = c("Study 1", "Study 2", "Study 3", "Study 4", "Study 5"))
## Error in jointmeta2(fits = c(JMfits2[1:3], JMfits[1:2]),
## longpar = c("time ", : Some of the joint model fits have differing
## association structures
```

The programmed error messages in the *jointmeta2* function are designed to help the user pool only parameters with comparable interpretations. In addition to flagging differing association structures, the function will also return an error if a mix of PH and AFT models have been used in the time-to-event sub-model of fits using the *JM* package. This functionality



should not replace the user considering what model fits are appropriate to pool, but aims to guard against the most common routes to pooling dissimilar parameters.

Currently, the *jointmeta2* function can only take fits produced by either the *joiner* or the *JM* packages. However, plans for future extensions of the R package include extending the function to take fits from additional packages such as single study univariate analyses conducted using the *joinerML* [181] and *joinerMeta* packages. Proposed further work on this package is discussed in Section 8.3.2.

### 4.3.2 Methods for One-Stage Meta-Analyses

#### 4.3.2.1 Function to fit One-Stage Models

The function *jointmeta1* fits a one-stage model to a multi-study joint data object in R of formats discussed in Chapter 3. The function has the following syntax:

```
jointmeta1(data, long.formula, long.rand.ind, long.rand.stud = NULL,
            sharingstrct = c("randprop", "randsep", "value", "slope",
                             "valandslope"), surv.formula, gpt, lgpt, max.it,
            tol, study.name, strat = F, longsep = F, survsep = F, bootrun = F
            print.detail = F)
```

The arguments of this function are as follows:

- ***data*** - a *jointdata* object containing the variables named in the model formulae.
- ***Long.formula*** - the formula for the longitudinal sub-model
- ***Long.rand.ind*** - the names of variables to assign individual level random effects to
- ***Long.rand.stud*** - the names of variables to assign study level random effects to
- ***sharingstrct*** - association structure - currently must be set to *"randprop"*
- ***surv.formula*** - the formula for the time-to-event sub-model
- ***gpt*** - the number of quadrature points across which the integration with respect to the random effects will be performed. This defaults to *gpt = 5*
- ***lgpt*** - the number of quadrature points which the log-likelihood is evaluated over following a model fit. This defaults to *lgpt = 7*
- ***max.it*** - the maximum number of iterations of the EM algorithm that the function will perform. Defaults to *max.it = 350*, however more iterations could be required for large complex datasets.

- ***tol*** - the tolerance level used to determine convergence in the EM algorithm. Defaults to *tol = 0.001*.
- ***study.name*** - name of the variable in the baseline dataset in the specified *jointdata* object giving study membership
- ***strat*** - if *TRUE* then the survival sub-model contains a baseline hazard stratified by study. If *FALSE*, the baseline is common across studies
- ***Longsep*** - if *TRUE*, the results from a separate longitudinal model with the same specification as the joint model's longitudinal sub-model are returned, *FALSE* otherwise
- ***survsep*** - if *TRUE*, the results from a separate time-to-event model with the same specification as the joint model's time-to-event sub-model (excluding the association structure) are returned, *FALSE* otherwise
- ***bootrun*** - if *TRUE*, the log-likelihood for the model is not calculated
- ***print.detail*** - if *TRUE* details of the parameters at each EM algorithm iteration are printed to the console

The formulas for the longitudinal sub-model and the time-to-event sub-model, supplied to arguments *Long.formula* and *surv.formula* respectively, are expressed in the same way as definition of the sub-models in the *joiner* package. Specifically, the longitudinal formula should be expressed as the name of the longitudinal outcome, followed by the specification for the model to fit to the longitudinal outcome, for example *Long.formula = Y ~ 1 + time + treat*. Interaction terms or functions can be specified, for example (as used in the real data analysis in Chapter 6) a specification of *Long.formula = Y ~ 1 + time + treat\*study + exp(-3\*time)* is permitted, to give a sub-model with an exponential of three times the longitudinal time term as well as an interaction term between treatment assignment and study membership. The function currently does not automatically support the inclusion of splines in the longitudinal sub-model, and only allows models that can be expressed as a linear combination of terms.

The time-to-event sub-model should be expressed using the *Surv* function (see the *survival* [195] package). As an example, *surv.formula = Surv(survtime, cens) ~ treat* would fit a time-to-event sub-model containing a single fixed effect of treatment assignment to time-to-event data with survival time *survtime* and censoring variable *cens*. The time-to-event sub-model, as with the longitudinal sub-model, can contain fixed interaction terms, for

example `Surv(survtime, cens) ~ treat*study`. Currently, the function only allows a proportional hazards model with an unspecified baseline hazard to be fitted. However, the baseline hazard can be common across studies (by setting argument `strat = FALSE`), or it can be stratified by study allowing a different baseline hazard for each study in the meta-analysis (by setting `strat = TRUE`).

Zero mean normally distributed random effects can be included at either the individual level (level 2), or at both the study level (level 3) and the individual level. The number of random effects at each level is capped at three in the current package, both for ease of programming when writing the model fitting function, and also to prevent the joint model from becoming overly complex. Random effects at any level should only be assigned to variables that are also assigned fixed effects in the model. This is because the study level random effects quantify the difference between overall population effects (quantified by the fixed terms) and the effect observed in each study, and the individual level random effects quantify the difference between the study effect and the individual effect (given that between study variation is accounted for, if not they quantify the difference between the overall population effect and the effect observed in the individual). Note, fixed effects do not have to be assigned random effects.

Unlike Wulfsohn and Tsiatis [78], who absorb the fixed effects in their joint model into the mean of their random effects distribution, the models proposed in this research separate fixed and random effects, resulting in individual and study level random effects with zero means, and the necessity to only assign random effects to coefficients that have been assigned fixed effects in the model (an approach reminiscent of the mixed effect models proposed by Laird and Ware [29]). The aim of this research is the meta-analysis of multi-study joint data, as such the parameters of interest are the population or overall effects of covariates (as quantified by the fixed effects). As such, separation of parameters into fixed effects, and zero mean random effects, simplifies the reporting of the parameters of interest to MA in the output from this model fitting function, facilitating the researcher in extraction of the required results.

If random effects were permitted non-zero means, as in Wulfsohn and Tsiatis [78], then the association structure would be closer to the current value than the random effects only parameterization. The value shared between sub-models would quantify how the recorded value(s) for an individual for any covariates assigned random effects effect their risk of an

event. However, fixed effects not assigned random effects would not be shared between sub-models (in Wulfsohn and Tsiatis [78], no standalone fixed effects are included in their model, so this does not occur). Future work for this software package aims to allow the user to fit a joint MA models with current value association structures (see Chapter 8).

Individual level random effects are specified by supplying the name or names of variables to be assigned individual level random effects to argument *rand\_ind*. If a random intercept is to be included in the model, then the character string *"int"* should be included to argument *rand\_ind*. As an example, to specify a model with an individual level random intercept and slope, then *rand\_ind = c("int", "time")*, where *time* would be the name of the longitudinal time variable. If an individual level random intercept is not to be included in the model, then *"noint"* should be included in the argument, e.g. *rand\_ind = c("noint", "time")*. At least one individual level random effect must be included in the model to ensure that the sub-models are linked. If no individual level random effects, or more than three, are specified, then the function returns an error message.

Study level random effects are specified in a way similar to the individual level random effects, in that the name or names of variables to be assigned study level random effects are supplied to argument *rand\_stud*. If a study level random intercept is to be included in the model, then the name of the study membership variable should be supplied to *rand\_stud*, e.g. *rand\_stud = c("study", "treat")*, otherwise the study membership variable name should not be present in *rand\_stud*. Study level random effects do not have to be included in the model, this can be achieved by not including the argument in the function call, or setting *rand\_stud = NULL*. An error message will be returned if more than three study level random effects are specified.

Currently, the *jointmeta1* function only allows the sub-models to be linked through the proportional random effects only structure [30, 75] (see Section 5.1.2.1 for methodology behind code, and Section 3.3.2 for examples of the types of models it is possible to fit using this function). As such, argument *sharingstrct* must be set to *"randprop"*, otherwise an error message will be encountered. It is planned to expand the function to permit additional association structures to be permitted (see Section 8.3.2.2), as such the argument *sharingstrct* has been included in the function call to facilitate this future work

Integration across the random effects included in the model is accomplished through pseudo-adaptive Gaussian quadrature [196]; this is further discussed in Section 5.1.2.2. Arguments *gpt* and *Lgpt* control the number of quadrature points for joint model fitting and estimation of the log-likelihood respectively.

The joint model is fitted using an EM algorithm [197], which requires starting values for parameters to be supplied. These starting values are generated by fitting initial separate longitudinal and time-to-event models, with the same specification as the corresponding sub-models in the joint model (apart from the association structure in the time-to-event sub-model). By setting *Longsep* and *survsep* to *TRUE*, the initial longitudinal and time-to-event model fits respectively are returned. This is useful if the user wishes to compare the results from the joint model and separate longitudinal and time-to-event models.

The argument *tol* specifies the tolerance level used when testing if the EM algorithm has converged. If the maximum difference between parameters estimated at two consecutive iterations is less than the value of *tol*, convergence is declared and the parameter estimates are returned. The default tolerance value is *tol* = 0.001. The argument *max.it* specifies the maximum number of iterations that the EM algorithm will perform before non-convergence is declared (the default value is 350 iterations, more iterations may be required for complex data).

The *bootrun* argument is included for use by the *jointmetaSE* function (described below). When fitting models at each bootstrap iteration, calculating the log-likelihood of the model fit is of no interest; it is only necessary to generate the parameter estimates. As such, setting *bootrun* to *TRUE* means that the model will not calculate the log-likelihood during the model fit (speeding up the bootstrapping process). In addition messages will not be printed from the function during each bootstrap's model fit. In general, when a model is fitted using *jointmeta1*, a message is printed when the EM algorithm commences the joint model fit, and also when calculation of the log-likelihood begins. In addition, during calculation of the log-likelihood, a progress bar is printed showing the percentage progress of the calculation. This was implemented as with a multi-study joint dataset, the log-likelihood calculation can take some time, and a moving progress bar provides reassurance that the function has not halted.

The *print.detail* argument, if set to *TRUE*, causes the estimates of the model parameters to be printed at each iteration of the EM algorithm. This is beneficial if the user wishes to see the process of each parameter's convergence during the model fitting process.

#### 4.3.2.2 Output from *jointmeta1*

The *jointmeta1* function generates a *jointmeta1.object*. The help file describing the output can be accessed once the *joineRmeta* package is loaded by typing *?jointmeta1.object* into the R console. In brief, the object contains an element *coefficient* which contains the coefficient estimates from the model. These are split into *fixed*, *random* and *Latent*. The fixed coefficients are themselves split into those from the longitudinal and those from the time-to-event sub-model. The *random* portion contains components *random\_ind* and *random\_stud*, which provide estimates of the random effect nodes for each individual for each individual level random effect (in *random\_ind*), and for each study for each study level random effect (in *random\_stud*) respectively. If study level random effects have not been specified, *random\_stud* will not be present. The component *Latent* contains estimates of any association parameters.

The object also contains the variance of the measurement errors in element *sigma.e* and the covariance matrices for the random effects in *rand\_cov* (split into elements *D* for the individual level random effects and, if present, *A* for the study level random effects). The estimated baseline hazard function is provided in element *hazard*, which is a single vector if argument *strat* was set to *FALSE*, a list of vectors (one for each study) if *strat* = *TRUE*.

The log-likelihood is provided in element *LogLik* (with components *jointLhood* for the overall log-likelihood, *jointy* for log-likelihood attributable to the longitudinal process, and *jointtn* for log-likelihood attributable to the time-to-event process). The number of completed iterations is returned in *numIter*, whether convergence was achieved (*TRUE/FALSE*) in element *convergence*, and the specified association structure in *sharingstrct*. Element *sepests* contains the separate longitudinal and time-to-event model fits if requested, otherwise it contains a message stating the results were not requested, and element *sep.LogLik* contains the log-likelihood from the separate model fits. The data supplied to the one-stage function is returned in element *data* and the original function call is held in element *Call*. Finally the number of studies present in the data is reported in element *numstudies*,

the number of individuals within each study is staged in element *n.bystudy*, the number of longitudinal measurements in each study is stated in *nobs*, and the ids of any individuals excluded from analysis due to insufficient data to include them in the joint model are given in *missingids*.

#### 4.3.2.3 Example of one-stage MA model fit

An example of a one-stage MA model fit is discussed below. This one-stage fit involves a fixed intercept, time, treatment and study membership terms in the longitudinal sub-model, along with an individual level random intercept and time term, and a study level random treatment effect. The association structure is set to *"randprop"* (currently this is the only association structure option). The time-to-event sub-model contains a fixed treatment effect only, as well as a baseline hazard stratified by study. The study membership variable is identified as the variable named *"study"*. The example involves the data held in *simdat*, which has been transformed into the *jointdata* format giving object *jointdat* (see Section 4.4.2.1).

```
onestagefit <- jointmeta1(data = jointdat, long.formula = Y ~ 1 + time +  
  treat + study, long.rand.ind = c("int", "time"),  
  long.rand.stud = c("treat"),  
  sharingstrct = "randprop",  
  surv.formula = Surv(survtime, cens) ~ treat,  
  study.name = "study", strat = TRUE)
```

This example will be used during the rest of the chapter to demonstrate the functions applicable to one-stage model fits in the package.

#### 4.3.2.4 Functions to extract information and perform bootstraps

Various functions have been included in the package to allow users to extract information from the model fit. These are discussed and demonstrated below, using the above example *onestagefit*. The functions include package specific versions of core functions expected to be available in any model fitting packages (*fixef*, *formula*, *print*, *ranef*, *summary*, *confint*, *vcov*), as well as some not found in other packages (*jointmetaSE* and *rancov*). Some of these functions are applied to the results of the *jointmeta1* function (i.e. they expect a multi-study joint model fit to be supplied); these are identified by the presence of *.jointmeta1* in their names. Others act on the results of the bootstrapping function *jointmetaSE*; these are identified by the presence of *.jointmeta1SE* in their names. These portions of function names (*.jointmeta1*, *.jointmeta1SE*) identify that the function is

written for use with *jointmeta1* and *jointmeta1SE* objects, produced by functions contained in the *joineRmeta* package. Functions specific to the *joineRmeta* package do not contain the suffix *.jointmeta1* or *.jointmeta1SE*.

#### 4.3.2.4.1 The *fixef.jointmeta1* function

The *fixef.jointmeta1* function extracts fixed effects from the one-stage multi-study joint model fit. The function can be applied to the model fit simply through the code *fixef(modelfit)*. R automatically identifies the class of model fit (*jointmeta1.object*) and finds the correct version of *fixef* to use. This structure of code applies to various other functions described below. The function call of the *fixef* function is as follows:

```
fixef(object, type = c("Longitudinal", "Survival", "Latent"), ...)
```

In the above function call, a *jointmeta1.object* should be supplied to argument *object*. The *fixef* function allows the user to extract the fixed effect coefficients from the longitudinal sub-model, by specifying argument *type = "Longitudinal"*, or to extract the coefficients from the time-to-event sub-model by specifying *type = "Survival"*. The association parameter(s) can be extracted by setting *type = "Latent"*:

```
fixef(object = onestagefit, type = "Longitudinal")
## (Intercept)      time      treat1      study2      study3      study4
## 0.28666444 2.84428399 1.56430760 0.60266225 0.09199601 -0.65413670
##      study5
## -1.38384289

fixef(object = onestagefit, type = "Survival")
##      treat1
## 1.969607

fixef(object = onestagefit, type = "Latent")
## gamma_ind_0 gamma_stud_0
## 0.98582075 -0.07625637
```

#### 4.3.2.4.2 The *formula.jointmeta1* function

The *formula.jointmeta1* function allows the model formulae defining various parts of the fitted joint model to be extracted. This function has the following format:

```
formula(x, type = c("Longitudinal", "Survival", "Rand_ind", "Rand_stud"), ...)
```

In this function, a *jointmeta1.object* should be supplied to argument *x*. The component of the joint model to display the formula for should be supplied to argument *type*, i.e. to extract



the formula for the fixed portion of the longitudinal sub-model set `type = "Longitudinal"`, for the time-to-event sub-model set `type = "Survival"`, for the individual level random effects set `type = "Rand_ind"`, and for the study level random effects set `type = "Rand_stud"`. The outputs of applying this function to the example model fit are shown below:

```
formula(x = onestagefit, type = "Longitudinal")
## Y ~ 1 + time + treat + study
formula(x = onestagefit, type = "Survival")
## Surv(survtime, cens) ~ treat
formula(x = onestagefit, type = "Rand_ind")
## ~1 + time
formula(x = onestagefit, type = "Rand_stud")
## ~ treat1
```

#### 4.3.2.4.3 The `print.jointmeta1` function

The `print.jointmeta1` function prints basic information about the one-stage joint model fit supplied to argument `x`. The function takes the following format:

```
print(x, ...)
```

Applying the function to a model fit gives:

```
print(x = onestagefit)
##
## Call:
## jointmeta1(data = jointdat, long.formula = Y ~ 1 + time + treat +
##   study, long.rand.ind = c("int", "time"), long.rand.stud = c("treat"),
##   sharingstrct = "randprop", surv.formula = Surv(survtime,
##   cens) ~ treat, study.name = "study", strat = TRUE)
##
## Random effects joint meta model
## Data: jointdat
##
## Longitudinal sub-model fixed effects: Y ~ 1 + time + treat + study
##
## (Intercept) 0.28666444
## time        2.84428399
## treat1      1.56430760
## study2      0.60266225
## study3      0.09199601
## study4     -0.65413670
## study5     -1.38384289
##
## Time-to-event sub-model fixed effects: Surv(survtime, cens) ~ treat
```

```
## Strat: TRUE
##   treat1
## 1.969607
##
##
## Latent association:
## gamma_ind_0  0.98582075
## gamma_stud_0 -0.07625637
##
## Variance components:
##           Type      Name      Value
## 1 Individual level (Intercept) 0.9909671
## 2                               time  1.240652
## 3           Study level      treat1 0.9171312
## 4           Residual              0.0041291
##
## Number of studies: 5
##
## Number of individuals per study:
##   1   2   3   4   5
## 500 500 500 500 500
##
## Number of longitudinal observations:
##   1   2   3   4   5
## 1422 1296 1346 1595 1752
```

The output of the function contains statements of the function call, the type of model fitted (a random effects joint meta model), the name of the data the function is fitted to, followed by a summary of the formulae and parameter estimates from each of the longitudinal and time-to-event sub-models. The association parameters are displayed under the "Latent Association" section. The "Variance Components" section shows the variance of each of the random effects at the individual and the study levels as well as the residual variance. Reports of the number of studies in the data, number of individuals per study, and number of longitudinal observations per study are then printed.

#### 4.3.2.4.4 The *rancov* function

The *rancov* function is a way for users to easily extract the estimated covariance matrices for the random effects at each level. The syntax of the function call is shown below:

```
rancov(fitted, type = c("individual", "study"))
```

The function expects a one-stage joint model fit of class *jointmeta1.object* to be supplied to argument *fitted*, whilst the level of the random effects for which to extract the covariance matrix needs to be supplied to argument *type*, (*type* = "individual" requests the covariance matrix for the individual level random effects, whilst *type* = "study" prints the covariance matrix for the study level random effects).

```

rancov(fitted = onestagefit, type = "individual")

##           (Intercept)      time
## (Intercept)  0.9909671 0.4800569
## time        0.4800569 1.2406520

rancov(fitted = onestagefit, type = "study")

##           treat1
## treat1 0.9171312

```

#### 4.3.2.4.5 The *ranef.jointmeta1* function

The *ranef.jointmeta1* function extracts the estimated nodes of the random effects at each level. The function call has syntax:

```

ranef(object, type = c("individual", "study"), ...)

```

In the function, a *jointmeta1.object* is to be supplied to argument *object*, whilst the level of random effects to display the estimated nodes for should be supplied to argument *type*. Setting *type = "individual"* returns a list with number of elements equal to the number of studies in the dataset. Each element of the returned list is a matrix with number of columns equal to the number of individual level random effects (two in the example), and number of rows equal to the number of individuals in the study. Setting *type = "study"* returns a matrix with number of columns equal to the number of study level random effects, and number of rows equal to the number of studies in the dataset.

An example is provided in the code below (with output requested for just the first five individuals in the first study; to print all results use *ranef(object = onestagefit, type = "individual")*).

```

ranef(object = onestagefit, type = "individual")[[1]][1:5,]

##      b2_(Intercept)    b2_time
## 464      -0.4541902  -0.2101875
## 104       2.1659179   1.0059201
## 479       2.2555976   1.0493788
## 448       1.0146335   0.4695370
## 319       0.6621034   0.3358462

ranef(object = onestagefit, type = "study")

##      b3_treat1
## 1  0.3515943
## 2  0.8482005
## 3 -1.4389932
## 4  1.0234457
## 5 -0.7843412

```

In the above example output, the results have columns labelled as random effects (with those acting at the individual level or level 2 of the data labelled *b2\_* followed by the name of the respective individual level random effect, and those acting at the study level or level 3 of the data labelled *b3\_* followed by the name of the respective study level random effect). The rows are labelled with the identifications of the units the random effect estimates relate to (if *type* = *"study"* the rows are labelled with the names of the studies in the dataset, whilst if *type* = *"individual"* the rows are labelled with the individual specific IDs).

#### 4.3.2.4.6 The *summary.jointmeta1* function

The *summary.jointmeta1* function is similar to the *print* function, and has format:

```
summary(object, variance = TRUE, ...)
```

In the above function, a one-stage joint model fit of class *jointmeta1.object* is supplied to argument *object*. The argument *variance* specifies whether the values for the variances (*variance* = *TRUE*) or for the standard deviations are shown (*variance* = *FALSE*). The result of applying *summary* to the example fit is shown below:

```
summary(object = onestagefit)

##
## Call:
## jointmeta1(data = jointdat, long.formula = Y ~ 1 + time + treat +
##   study, long.rand.ind = c("int", "time"), long.rand.stud = c("treat"),
##   sharingstrct = "randprop", surv.formula = Surv(survtime,
##   cens) ~ treat, study.name = "study", strat = TRUE)
##
## Random effects joint meta model
## Data: jointdat
## Log-likelihood: -14722.25
## AIC: 29472.49
##
## Longitudinal sub-model fixed effects: Y ~ 1 + time + treat + study
## (Intercept) 0.28666444
## time        2.84428399
## treat1      1.56430760
## study2      0.60266225
## study3      0.09199601
## study4     -0.65413670
## study5     -1.38384289
##
## Time-to-event sub-model fixed effects: Surv(survtime, cens) ~ treat
## Strat: TRUE
##   treat1
## 1.969607
##
## Latent association:
## gamma_ind_0 0.98582075
```

```
## gamma_stud_0 -0.07625637
##
## Variance components:
##      Type      Name      Value
## 1 Individual level (Intercept) 0.9909671
## 2      time      1.240652
## 3 Study level      treat1 0.9171312
## 4      Residual      0.0041291
##
## Convergence at iteration: 13
##
## Number of studies: 5
##
## Number of individuals per study:
##   1   2   3   4   5
## 500 500 500 500 500
##
## Number of longitudinal observations:
##   1   2   3   4   5
## 1422 1296 1346 1595 1752
```

The output of the function is similar to that obtained by `print(onestagefit)`, but slightly more information is returned. For example, a statement of the log-likelihood and the AIC is included along with the number of EM algorithm iterations conducted before convergence was achieved.

#### 4.3.2.4.7 The *jointmetaSE* function

The model fitted using the *jointmeta1* function includes an unspecified baseline hazard in the time-to-event sub-model. Hsieh et al [89] recommend that standard errors should be calculated through a bootstrapping procedure to avoid underestimation. A bootstrapping procedure is provided in the *jointmetaSE* function, which has the following function call:

```
jointmetaSE(fitted, n.boot, gpt, max.it, tol, print.detail = FALSE,
             overalleffects = NULL)
```

Here, a one-stage joint model fit of class *jointmeta1.object* should be supplied to argument *fitted*. The number of bootstraps to complete should be supplied to *n.boot*, and the number of quadrature points to integrate the random effects over during model fitting should be supplied to *gpt* (this will default to 5). The maximum number of iterations to complete would be supplied to *max.it* (this defaults to 500), and the tolerance used to test convergence should be supplied to *tol* (this defaults to 0.001). If argument *print.detail* is set to *TRUE*, the parameter estimates at each iteration of each bootstrap will be printed. If this is set to *FALSE* then instead a progress bar showing completion of all bootstraps will be printed.

The code to apply the bootstrapping function to a multi-study joint model fit is:

```
onestagefitSE <- jointmetaSE(fitted = onestagefit, n.boot = 200)
```

The *jointmetaSE* function outputs a *jointmeta1SE.object*. The help file for this class of object can be displayed by typing *?jointmetaSE.object* into the R console when the *joinermeta* package is loaded.

The structure of a *jointmeta1SE* object can be viewed using code such as *str(onestagefitSE)*, which shows a list of three objects, labelled *results*, *covmat*, and *bootstraps*. The object *bootstraps* contains the parameter estimates from every model fit performed during the bootstrap process, and can be accessed by typing *onestagefitSE\$bootstraps*. The object *results* gives the parameter estimates, standard errors and confidence intervals in a neatly arranged table, and can be obtained by typing *onestagefitSE\$results* into the console. The covariance matrix for the model parameters is accessible by running code *onestagefitSE\$covmat*. The bootstrapping procedure will only return confidence intervals if at least 100 bootstraps have been completed.

A range of functions to work with an object of class *jointmeta1SE* are included in the package, to allow easy extraction of information, these are discussed in the following sections.

#### 4.3.2.4.8 The *print.jointmeta1SE* function

The print function *print.jointmeta1SE* has function syntax:

```
print(x, ...)
```

The results of the *jointmeta1SE* function, i.e. a *jointmetaSE.object* should be supplied to argument *x*. This will extract and return the *results* element of a *jointmetaSE.object*. When applied to the example, the results of the bootstrapping procedure are returned:

```
print(onestagefitSE)
```

##	Component	Parameter	Estimate	SE	95%Lower	95%Upper
## 1	Longitudinal	(Intercept)	0.2867	0.0654	0.1578	0.4073
## 2		time	2.8443	0.0288	2.787	2.9034
## 3		treat1	1.5643	0.0379	1.4882	1.6386
## 4		study2	0.6027	0.0831	0.4409	0.7623
## 5		study3	0.092	0.0865	-0.0621	0.2503
## 6		study4	-0.6541	0.089	-0.8352	-0.4991
## 7		study5	-1.3838	0.0874	-1.5591	-1.2188
## 8	Survival	treat1	1.9696	0.0727	1.8322	2.134
## 9	Association	gamma_ind_0	0.9858	0.0227	0.9413	1.0299

```
## 10      gamma_stud_0 -0.0763 0.0868 -0.2533 0.0646
## 11      Variance b2_(Intercept) 0.991 0.0295 0.924 1.036
## 12      b2_time 1.2407 0.0405 1.1603 1.3139
## 13      b3_treat1 0.9171 0.0718 0.7855 1.0634
## 14      Residual 0.0041 1e-04 0.0039 0.0044
```

#### 4.3.2.4.9 The *confint.jointmeta1SE* function

The package also contains *confint.jointmeta1SE* which extracts only the confidence intervals from the results of the bootstrapping process, and has the following formulation.

```
confint(object, parm = NULL, level = 0.95, ...)
```

In the above function, a *jointmetaSE.object* should be supplied to argument *object*, *parm* is a vector giving the parameter names or locations in the list of model parameters for which confidence intervals should be returned, and the level of the confidence intervals should be supplied to argument *Level* (this defaults to 0.95 for 95% confidence intervals). This function was included to mirror the common practice of model fitting packages to include a function to extract only parameter estimates and confidence intervals from model fitting results, and also to allow the user to calculate confidence intervals at a level other than 95%.

This function can be demonstrated using the example bootstraps object *onestagefitSE*. The first example requests 95% confidence intervals for all the parameters included in the model (as argument *parm* is not specified, all results are returned, and as *Level* is not specified it defaults to 0.95).

```
confint(object = onestagefitSE)

##      Component      Parameter Estimate 95%Lower 95%Upper
## 1 Longitudinal (Intercept) 0.2867 0.1578 0.4073
## 2      time 2.8443 2.787 2.9034
## 3      treat1 1.5643 1.4882 1.6386
## 4      study2 0.6027 0.4409 0.7623
## 5      study3 0.092 -0.0621 0.2503
## 6      study4 -0.6541 -0.8352 -0.4991
## 7      study5 -1.3838 -1.5591 -1.2188
## 8      Survival      treat1 1.9696 1.8322 2.134
## 9 Association      gamma_ind_0 0.9858 0.9413 1.0299
## 10      gamma_stud_0 -0.0763 -0.2533 0.0646
## 11      Variance b2_(Intercept) 0.991 0.924 1.036
## 12      b2_time 1.2407 1.1603 1.3139
## 13      b3_treat1 0.9171 0.7855 1.0634
## 14      Residual 0.0041 0.0039 0.0044
```

Confidence intervals can be requested just for one set of parameters (e.g. just the treatment effects), and can be requested at a different level (e.g. 99%) using the following code:

```
confint(object = onestagefitSE, parm = "treat1", level = 0.99)
##      Component Parameter Estimate 99%Lower 99%Upper
## 3 Longitudinal   treat1   1.5643   1.4510   1.6509
## 8      Survival   treat1   1.9696   1.7931   2.1694
```

4.3.2.4.10 The `vcov.jointmeta1SE` function

Finally, the `vcov.jointmeta1SE` function automatically extracts the covariance matrix from an object of class `jointmeta1SE` when it is supplied to argument `object`, and has syntax:

```
vcov(object, ...)
```

As an example, running the code `vcov(object = onestagefitSE)` will extract the covariance matrix for the model parameters.

## 4.4 Exploratory Functions

A range of tools designed for the exploratory analysis of multi-study joint data are available in the `joineRmeta` package. These are divided into three main sections; functions for Data Simulation (Section 4.4.1), for Data Preparation (Section 4.4.2), and for Data Visualisation (Section 4.4.3). Functions described in this section are designed to be used prior to models being fitted to the data, and aim to help the users generate multi-study joint data to investigate models under controlled conditions, to prepare data into the format required by the model fitting functions, or to produce visual summaries of the data to aid with model choice (e.g. for use in Preliminary work before a two-stage MA, see Section 3.2.1).

### 4.4.1 Data Simulation

Simulation studies are a powerful tool to investigate model behavior under specified conditions. At the start of this thesis, functions to simulate joint data in a single study case were available, however multi-study data simulation functions for joint longitudinal and time-to-event data that allowed control over study level variation through study level random effects were not available. As such, the data simulation function included in the `joineR` package [80] has been expanded to permit simulation of such data.

The data simulation function `simjointmeta` allows the user to generate joint longitudinal and time-to-event data for multiple studies simultaneously. The function and its arguments are shown below, with the default settings for the arguments described:



```
simjointmeta(k = 5, n = rep(500, 5), sepassoc = FALSE, ntms = 5,
  longmeasuretimes = c(0, 1, 2, 3, 4), beta1 = c(1, 1, 1),
  beta2 = 1, rand_ind = c("intslope", "int"),
  rand_stud = c("int", "inttreat", "treat", NULL), gamma_ind = 1,
  gamma_stud = NULL, sigb_ind, sigb_stud = NULL, vare = 0.01,
  theta0 = -3, theta1 = 1, censoring = TRUE, censlam = exp(-3),
  truncation = FALSE, trunctime = max(longmeasuretimes))
```

The function contains a range of arguments which are defined below.

- ***k*** - the number of studies to be simulated
- ***n*** - a vector (of length equal to *k*) specifying the number of individuals to be simulated per study
- ***sepassoc*** - a *TRUE/FALSE* value specifying whether the data should be generated using separate association parameters per random effect (*TRUE*), or whether the association parameter is common across random effects acting at the same level (*FALSE*), see Section 1.6.2 for an introduction to association structures.
- ***ntms*** - the number of longitudinal time-points at which a measurement is possible (this should be equal to the number of *Longmeasuretimes* specified, see next argument).
- ***Longmeasuretimes*** - a vector of longitudinal measurement times. If this is not specified in the function call, *Longmeasuretimes* will be set to equal integer values from 0 to *ntms-1*.
- ***beta1*** - a vector containing the longitudinal fixed effect coefficients in order intercept, treatment effect, time
- ***beta2*** - a vector containing the time-to-event fixed effect coefficient for treatment effect
- ***rand\_ind*** - the specification of individual level (level 2) random effects (individual level random effects must be present in the model the data is simulated under, to ensure random effects are present to link the sub-models). This argument can take the following values:
  - ***"intslope"*** - a model with individual level random intercept and slope (time) terms
  - ***"int"*** - a model with individual level random intercept
- ***rand\_stud*** - the specification of study level (level 3) random effects. Inclusion of study level random effects in the model the data is simulated under is optional. This argument can take the following values:

- *NULL* - no study level random effects (equivalent to not specifying the argument in the function call)
- *"int"* - a model with study level random intercept
- *"treat"* - a model with study level random treatment effect
- *"inttreat"* - a model with study level random intercept and random treatment effect terms.
- *gamma\_ind* - the association parameter(s) linked to the individual level random effects.
- *gamma\_stud* - the association parameter(s) linked to the study level random effects (if included)
- *sigb\_ind* - the covariance matrix for the individual level random effects
- *sigb\_stud* - the covariance matrix for the study level random effects (if included)
- *vare* - the variance of the longitudinal measurement error term
- *theta0* - a parameter used to control the distribution of the event times (the specification of the individual level random effects dictates the distribution of the event times)
- *theta1* – a parameter used to control the distribution of the event times
- *censoring* - an argument indicating whether the event times should be subject to censoring (*TRUE*) or not (*FALSE*)
- *cenlam* - a parameter controlling the exponentially distributed censoring times
- *truncation* - an argument indicating if the simulated survival times (the minimum of the event and censoring time if censoring is possible, the event time otherwise) are truncated at a certain value (*TRUE*) or not (*FALSE*)
- *trunc time* - if *truncation* is true, the time at which the survival times are to be truncated

As mentioned, study level random effects are specified through argument *rand\_stud*. Under this function, it is possible to simulate data with variation between studies in treatment effect, or in the intercept, or both, or with no variation between studies due to study level random effects. As discussed in Section 3.3.2, inclusion of a study level random intercept introduces between study variation in the longitudinal trajectory intercept and in the log hazard ratio for risk of an event between data from two studies. Inclusion of a study level random treatment

effect introduces between study variation in the differences in the longitudinal trajectories between treatment groups, and in the log hazard ratio for risk of an event between treatment groups. If no study level random effects are specified, data simulated will be more similar between studies; however data will not be identical due to the use of realizations from random variables during the simulation process e.g. when generating error terms or the survival times.

The function allows data to be generated with either separate association (*sepassoc* = *TRUE*) or common association (*sepassoc* = *FALSE*) parameters for the random effects at a given level. For example, if an individual level random intercept and time term, and a study level random intercept and treatment term were specified, setting *sepassoc* = *TRUE* causes the  $W_{2ki}(t)$  term inserted into the time-to-event sub-model to have structure  $W_{2ki}(t) = \alpha_0^{(2)} b_{0ki}^{(2)} + \alpha_1^{(2)} b_{1ki}^{(2)} T_{Ski} + \alpha_0^{(3)} b_k^{(3)} + \alpha_1^{(3)} b_{1k}^{(3)} treat_{ki}$ . However, setting *sepassoc* = *FALSE* gives  $W_{2ki}(t)$  with specification  $W_{2ki}(t) = \alpha^{(2)} (b_{0ki}^{(2)} + b_{1ki}^{(2)} T_{Ski}) + \alpha^{(3)} (b_k^{(3)} + b_{1k}^{(3)} treat_{ki})$ .

If *sepassoc* is set to *TRUE*, then the number of association parameters supplied to *gamma\_ind* should be the same as the specified number of individual level random effects in the model the data is generated under. For example, if *rand\_ind* is set to "*intslope*", then the function would expect two association parameters be supplied to *gamma\_ind*. Similarly, if study level random effects are specified, then the number of association parameters supplied to *gamma\_stud* and the number of study level random effects should be equal.

Between study heterogeneity can also be introduced to the data through the association parameters. If the user wishes to specify different association parameters for each study, then the association parameter arguments *gamma\_ind* and *gamma\_stud* can be supplied as lists of length equal to the number of studies. Each element of the list would be either a single value if *sepassoc* = *FALSE* or a vector of values equal in length to the number of random effects acting at the relevant level if *sepassoc* = *TRUE*.

The arguments *theta0* and *theta1* control the distribution of the event times for individuals in the dataset. As with the association parameters, between study heterogeneity can be introduced by specifying different values for the arguments for each simulated study. If the parameters defining the distribution of the event times is to be identical between studies, *theta0* and *theta1* should be defined as single values in the function call. Otherwise, *theta0* and *theta1* should each be specified as a vector of values equal in length to the

number of studies to be simulated. Further details concerning the effect of arguments *theta0* and *theta1* are given in Section 5.2.

The argument *censLam* controls the distribution of the censoring times (if censoring is permitted in the data by setting *censoring* = *TRUE*). As with the parameters controlling the distribution of event times, the *censLam* argument can be supplied as a single value, or as a vector of value of length equal to the number of studies to be simulated. This allows the distribution of censoring times to be the same or varying across studies, allowing another way for users to introduce between study heterogeneity to their data.

The methodology used by this function to simulate multi-study joint data is detailed in Section 5.2.

An example of the data simulation function and its output is presented below. When the data is simulated, the percentage event rate is printed to the R console, but it is also saved as part of the output. To simulate 5 studies each containing 500 individuals the following code could be run:

```
exampledat1 <- simjointmeta(k = 5, n = rep(500, 5), sepassoc = FALSE,
  ntms = 5, longmeasuretimes = c(0, 1, 2, 3, 4),
  beta1 = c(1, 2, 3), beta2 = 1,
  rand_ind = "intslope", rand_stud = NULL,
  gamma_ind = 1,
  sigb_ind = matrix(c(1, 0.5, 0.5, 1.5), nrow = 2),
  vare = 0.01, theta0 = -3, theta1 = 1,
  censoring = TRUE, censlam = exp(-3),
  truncation = FALSE,
  trunctime = max(longmeasuretimes))

## 72.2 % experienced event
## 75 % experienced event
## 75.8 % experienced event
## 71.6 % experienced event
## 79.2 % experienced event
```

The exact structure of the simulated data can be checked by running the code *str(exampledat1)* in the R console. The data produced by the simulation function is a list of 3 elements, *Longitudinal* which is a list of study specific long format datasets, *survival* which is a list of study specific event datasets, and *percentevent* which is a list of values stating the event rate in each simulated study. This simulated data can be changed into a *jointdata* object using the function *tojointdata*.

In each case, for each study, the longitudinal dataset contains an identification variable *id* for each individual, the continuous longitudinal outcome *Y*, the continuous longitudinal time variable *time*, a study membership variable *study*, an intercept (which always takes value 1) *intercept*, a binary (0/1) treatment variable *treat*, and a duplicate of the longitudinal time variable *Ltime* (this variable is duplicated due to the data simulation function, and could be discarded).

The survival datasets each contain an identification variable *id*, the survival time for each individual *survtime*, a censoring indicator *cens* (with 1 indicating an event, and 0 otherwise), a study membership variable *study*, and a binary treatment group indicator *treat*. Identical values in the *id* variables between the longitudinal and time-to-event datasets for a particular study identify the same individual.

The final element of the list, *percentevent*, contains a list of the event rate in each study, with the first element of the list giving the event rate in the first study included in the dataset, and so on.

#### 4.4.2 Data Preparation

##### 4.4.2.1 Data formatting

Like the *joiner* package, *joinerMeta* expects joint data to be supplied as an object of class *jointdata*. This object is defined in the *joiner* package [80], and contains a list of six elements: namely a vector of the individual identification or *id* variable, a longitudinal dataset, a time-to-event or survival dataset, a baseline dataset, the name of the longitudinal time variable, and the name of the individual identification or *id* variable. The object has to contain at least either a longitudinal or a survival dataset. Given the expected multi-study nature of the data, the package contains function *tojointdata* to aid users reshape data into the required *jointdata* format.

Multi-study joint modelling data can be made available to researchers in a range of formats. Separate datasets could be provided for each study in the meta-analysis. Longitudinal and time-to-event data could be supplied in the same dataset, in either long format (one line per longitudinal measurement, potentially multiple lines per individual), or wide format (one line per individual, longitudinal measurements supplied in multiple columns). Longitudinal and

time-to-event data could also be provided in separate datasets. Additional datasets containing baseline information could also be available. The *tojointdata* function is designed to take data in various formats, and output a *jointdata* object. It has the following syntax:

```
tojointdata(dataset = NULL, longitudinal = NULL, survival = NULL,  
            baseline = NULL, id, longoutcome, timevarying = NULL, survtime, cens,  
            time = NULL, longtimes = NULL)
```

The arguments of the *tojointdata* function are defined as follows:

- **dataset** - a dataset or list of datasets
- **Longitudinal** - a dataset or list of datasets in long format containing the longitudinal outcome and any time varying covariates (baseline information can also be included)
- **survival** - a dataset or list of datasets in wide format containing the survival time and censoring variable (baseline information can also be included)
- **baseline** - a dataset or list of datasets in wide format containing any baseline information. This does not have to be supplied
- **id** - the name of the identification variable
- **Longoutcome** - the name of the longitudinal outcome variable
- **timevarying** - a vector of the names of any time varying covariates in the dataset
- **survtime** - the name of the survival time variable
- **cens** - the name of the censoring indicator variable
- **time** - if data is supplied in wide format, the name to assign to the longitudinal time variable produced by transforming the longitudinal data from wide to long format. If data is supplied in long format, the name of the longitudinal time variable
- **Longtimes** - if wide data, labels identifying the time points that the time varying variables relate to

It is assumed that data from different studies can be supplied to the *tojointdata* function in the same format, e.g. all studies provide datasets in wide format containing both longitudinal and time-to-event data. If this is not the case, reformatting the data is straightforward using packages such as *reshape2* [198], *tidyr* [199], or using functions available in the base R code such as *reshape*.

The argument *dataset* can be supplied as a single dataset containing all longitudinal and time-to-event data from all studies, or can be supplied as a list of study specific datasets containing both longitudinal and time-to-event information. These datasets can be in wide or long format. It is expected that either the argument *dataset* is specified, or the arguments *Longitudinal* and *survival* (with *baseline* if additional baseline data is available) are specified.

If data is supplied in wide format, the argument *Longtimes* should be supplied as a vector of labels identifying the time points that the time-varying variables relate to. For example if there are four time points, and a time varying variable *x* in the wide dataset labelled *x\_1*, *x\_2*, *x\_3*, and *x\_4*, then *Longtimes=c(1,2,3,4)* would be supplied to the function.

An example of using the *tojointdata* function applied to simulated dataset *simdat* to obtain a *jointdata* object (with the structure of the resulting output printed) is displayed below:

```
jointdat<-tojointdata(longitudinal = simdat$longitudinal,
  survival = simdat$survival, id = "id", longoutcome = "Y",
  timevarying = c("time", "ltime"), survtime = "survtime", cens = "cens",
  time = "time")
jointdat$baseline$study<-as.factor(jointdat$baseline$study)
jointdat$baseline$treat<-as.factor(jointdat$baseline$treat)
str(jointdat)

## List of 6
## $ subject      : num [1:2500] 1 2 3 4 5 6 7 8 9 10 ...
## $ longitudinal:'data.frame': 7411 obs. of 4 variables:
## ..$ id       : num [1:7411] 1 2 3 4 5 5 5 5 5 6 ...
## ..$ Y        : num [1:7411] 0.847 3.04 -0.252 1.976 -1.23 ...
## ..$ time     : num [1:7411] 0 0 0 0 0 1 2 3 4 0 ...
## ..$ ltime    : num [1:7411] 0 0 0 0 0 1 2 3 4 0 ...
## $ survival     :'data.frame': 2500 obs. of 3 variables:
## ..$ id        : num [1:2500] 1 2 3 4 5 6 7 8 9 10 ...
## ..$ survtime  : num [1:2500] 0.346 0.595 0.489 0.323 6 ...
## ..$ cens      : num [1:2500] 1 0 0 1 0 0 0 1 1 1 ...
## $ baseline     :'data.frame': 2500 obs. of 4 variables:
## ..$ id        : num [1:2500] 1 2 3 4 5 6 7 8 9 10 ...
## ..$ study     : Factor w/ 5 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
## ..$ treat     : Factor w/ 2 levels "0","1": 1 2 1 2 1 1 2 2 1 2 ...
## ..$ intercept: num [1:2500] 1 1 1 1 1 1 1 1 1 1 ...
## $ time.col     : chr "time"
## $ subj.col     : chr "id"
## - attr(*, "class")= chr "jointdata"
```

#### 4.4.2.2 Data cleaning

It is possible that longitudinal measurements are available for individuals after their survival time, if the event in question is non-terminal, such as time to next exacerbation in asthma

datasets. However, longitudinal data measured after the event of interest are not included in the joint model. As such, a function *removeafter* is included in the *joinerMeta* package, designed to remove any longitudinal measurements measured after the individual's survival time. This function has the following syntax:

```
removeafter(data, longitudinal, survival, id, time)
```

- **data** - a *jointdata* object
- **longitudinal** - the name of the longitudinal variable
- **survival** - the name of the survival time variable
- **id** - the name of the individual identification or id variable for the data
- **time** - the name of the longitudinal time variable

The *removeafter* function must be supplied with data in the *jointdata* format, and also returns output in *jointdata* format. The *removeafter* function is demonstrated using the example dataset *simdat3* from the package, which contains longitudinal measurements recorded after the survival time. Firstly the dataset is reformatted into a *jointdata* object.

```
jointdat3<-tojointdata(longitudinal = simdat3$longitudinal,  
  survival = simdat3$survival, id = "id", longoutcome = "Y",  
  timevarying = c("time","ltime"), survtime = "survtime", cens = "cens",  
  time = "time")
```

Once *simdat3* is in *jointdata* format, the *removeafter* function can be used to remove any longitudinal information recorded after an individual's event. When this function is run in the R console, a progress bar is printed to display the progress of the function.

```
jointdat3.1<-removeafter(data = jointdat3, longitudinal = "Y",  
  survival = "survtime", id = "id", time = "time")
```

If the structures of the original *jointdata* object containing all longitudinal measurements (*jointdat3*) and the *jointdata* object with longitudinal measurements recorded after the survival time removed (*jointdat3.1*) are compared, the removal of some longitudinal data (any measurements recorded after the individual in question's survival time) can be observed (12500 measurements versus 5846 measurements).



#### 4.4.3 Data Visualization

Two data visualization functions are currently included in the *joinerMeta* package. These functions produce study specific plots of the longitudinal and the time-to-event data, and arrange them neatly to display all studies on the same plot. These plotting functions use the flexible *ggplot2* package [200] to create plots tailored to multi-study joint data.

##### 4.4.3.1 Study specific plotting function

The *jointmetaplot* function is designed to produce the plots described in Chapter 3, to enable an assessment of the potential correlation between the longitudinal and time-to-event outcomes, and the suitability of different models for the longitudinal and time-to-event sub-model. The function produces study specific graphs, and has the following function call:

```
jointmetaplot(dataset, study, longoutcome, longtime, survtime, cens, id,
  smoother = FALSE, studynames = NULL, type = c("Longitudinal", "Event",
  "Both"), eventby = NULL, eventconfint = FALSE)
```

- **dataset** - a *jointdata* object
- **study** - the name of the study membership variable
- **Longoutcome** - the name of the longitudinal outcome variable
- **Longtime** - the name of the longitudinal time variable
- **survtime** - the name of the survival time variable
- **cens** - the name of the censoring indicator variable
- **id** - the name of the ID variable
- **smoother** - if *TRUE* a smoother is included over the longitudinal trajectory plots for each study, *FALSE* otherwise
- **studynames** - a vector of labels for the study specific plots, e.g. the first element will be the label for any plot produced from data from the first study listed in the dataset
- **type** - argument to select the type of plot to produce
- **eventby** - optional argument giving the name of a grouping variable to stratify the Kaplan-Meier curves by if plots of the time-to-event data are requested
- **eventconfint** - if *TRUE*, confidence intervals will be displayed for the time-to-event plots, defaults to *FALSE*

This function is designed to easily permit users to create identical plots of the longitudinal and or the time-to-event data in a multi-study joint dataset. The first arguments simply define the dataset to be used, with key information such as the names of the outcomes and individual identification or id variables. The function also contains a range of arguments to control the appearance of the plot.

The argument *type* allows the user to specify whether just plots of the longitudinal trajectories are to be produced (*type* = *"Longitudinal"*), whether just Kaplan-Meier plots of the time-to-event data are to be returned (*type* = *"Event"*), or whether the user wants the function to return both plot types (*type* = *"Both"*).

The longitudinal plots produced by the function show the longitudinal trajectories for all individuals within a dataset. If the studies include many individuals, the trajectories can become heavily overlaid, and so the behaviour of the trajectories can be difficult to distinguish. A way to assess the overall behaviour of the trajectories is to add a smoother over the trajectories. By setting the argument *smoother* = *TRUE* inserts a loess smoother over the individual trajectories.

The time-to-event plots produced by the function show the Kaplan-Meier [177] plots for each study. Various options exist for the time-to-event plots. It is often important to compare the curves between different groups. In this case, the name of a categorical grouping variable, such as treatment assignment, can be supplied to argument *eventby* to produce separate curves for each group. Another option is to add confidence intervals to the survival curve; this can be achieved by setting argument *eventconfint* = *TRUE*.

The function returns an object of class *jointplots*. If argument *type* is set to either *"Longitudinal"* or *"Both"* then the output will contain an element labelled *Longplots*, which is a list of *ggplot2* objects (see the *ggplot2* package [200]) plotting, for each study, the individual longitudinal trajectories. If argument *type* is set to either *"Event"* or *"Both"* then the output will contain an element labelled *eventplots*, which is a list of *ggplot2* objects plotting the survival probabilities for each study. Note, in each case if they are returned, element *Longplots* and element *eventplots* are lists of length equal to the number of studies in the dataset.

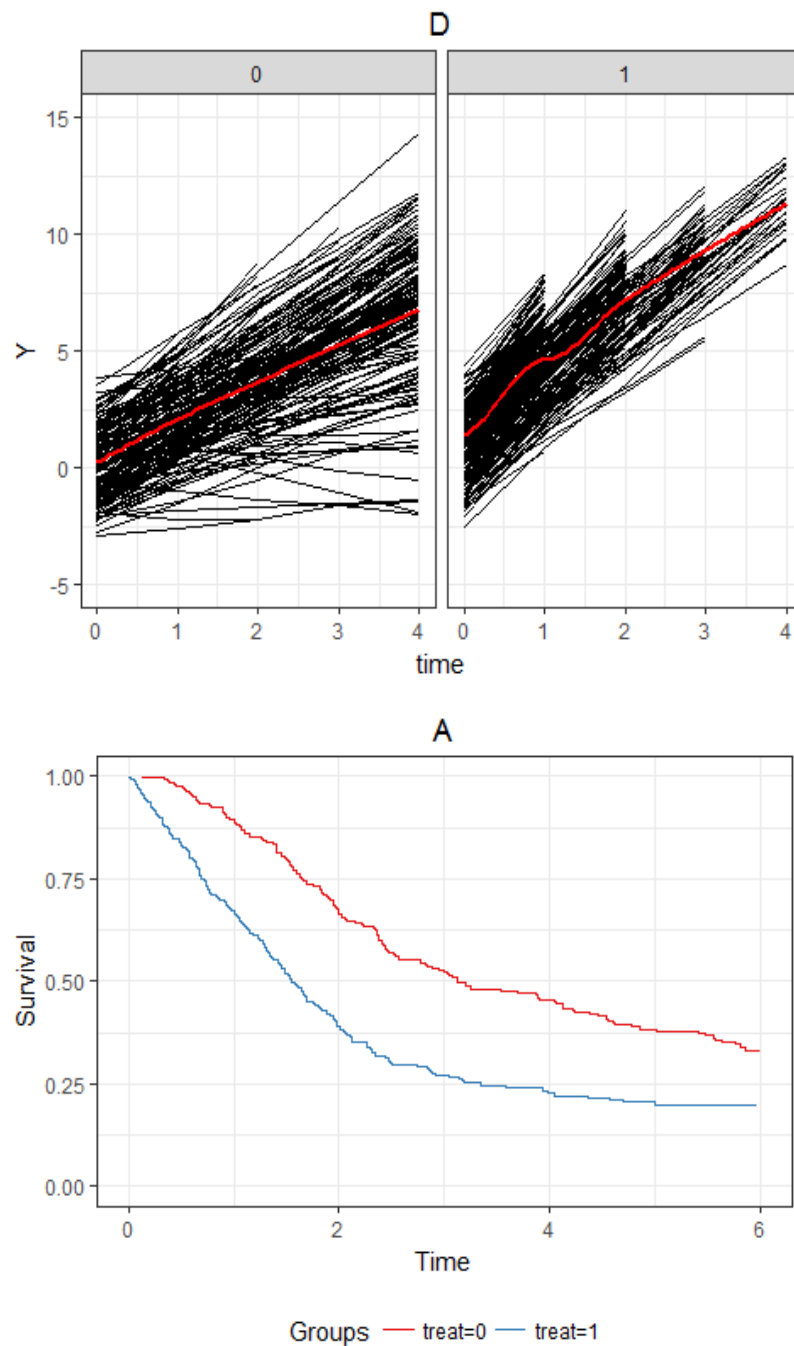
The elements of *Longplots* or *eventplots* are given the labels supplied in argument *studynames*. Individual plots can be extracted either by their placement in the list (e.g. *Longplots[[1]]* to extract the longitudinal plot for the first study), or can be extracted by name as specified in *studynames*, e.g. *Longplots\$studyname1*.

As an example, plots of the longitudinal trajectories and survival curves of an example dataset can be obtained, using the following code.

```
studypLOTS<-jointmetaplot(dataset = jointdat, study = "study",
  longoutcome = "Y", longtime = "time",
  survtime = "survtime", cens = "cens", id = "id",
  smoother = TRUE,
  studynames = c("A", "B", "C", "D", "E"),
  type = "Both", eventby = "treat",
  eventconfint = FALSE)
```

The plots produced can be examined using the following code, noting that plots can be extracted by name (first line), or through their placement in the list (second line).

```
studypLOTS$longplots$`studyplot.D`
studypLOTS$eventplots[[1]]
```



#### 4.4.3.2 Function to arrange study specific plots in grid

The *jointmetaplot* function returns study specific plots, however it is often difficult to compare multiple plots at once. As such, a function *jointmetaplotall* is provided in the *joinermeta* package to arrange the plots produced by the *jointmetaplot* function into

grid of graphs that can be output as a single image. The function call for this function is shown below:

```
jointmetaplotall(plotlist, ncol, nrow = NULL, top = NULL,  
  type = c("Longitudinal", "Event", "Both"))
```

- **plotlist** - the output produced by the *jointmetaplot* function
- **ncol** - number of columns in grid to output graphs in (must be specified)
- **nrow** - number of rows in grid to output graphs in (optional)
- **top** - a title for the grid plots
- **type** - as with the *jointmetaplot* function, types of plots to return

The output from the *jointmetaplot* function can be saved to an object in the R workspace, e.g. *studyplots* and supplied to the *plotlist* argument by *plotlist = studyplots*. The study specific graphs will be arranged in a grid, with common legend where appropriate e.g. if survival curves were stratified by a grouping variable.

An object of class *jointplotsall* will be returned by the function, which again is a list, the elements of which depend on the function argument *type*. If *type* is set to "Longitudinal" or "Both", then an element "Longall" will be present in the returned object, giving the study specific longitudinal trajectory graphs arranged in a single grid. If *type* is set to "Event" or "Both", then an element "eventsall" will be present, giving the study specific survival curves arranged in a single grid.

The dimensions of the grid are controlled by arguments *ncol* and *nrow*. Argument *ncol* has to be defined in the function call, whilst *nrow* is optional, and if missing will be calculated based on the number of study specific plots supplied to *plotlist* and the specified value of *ncol*. A common title can be supplied to any grid plots output by the function using argument *top*.

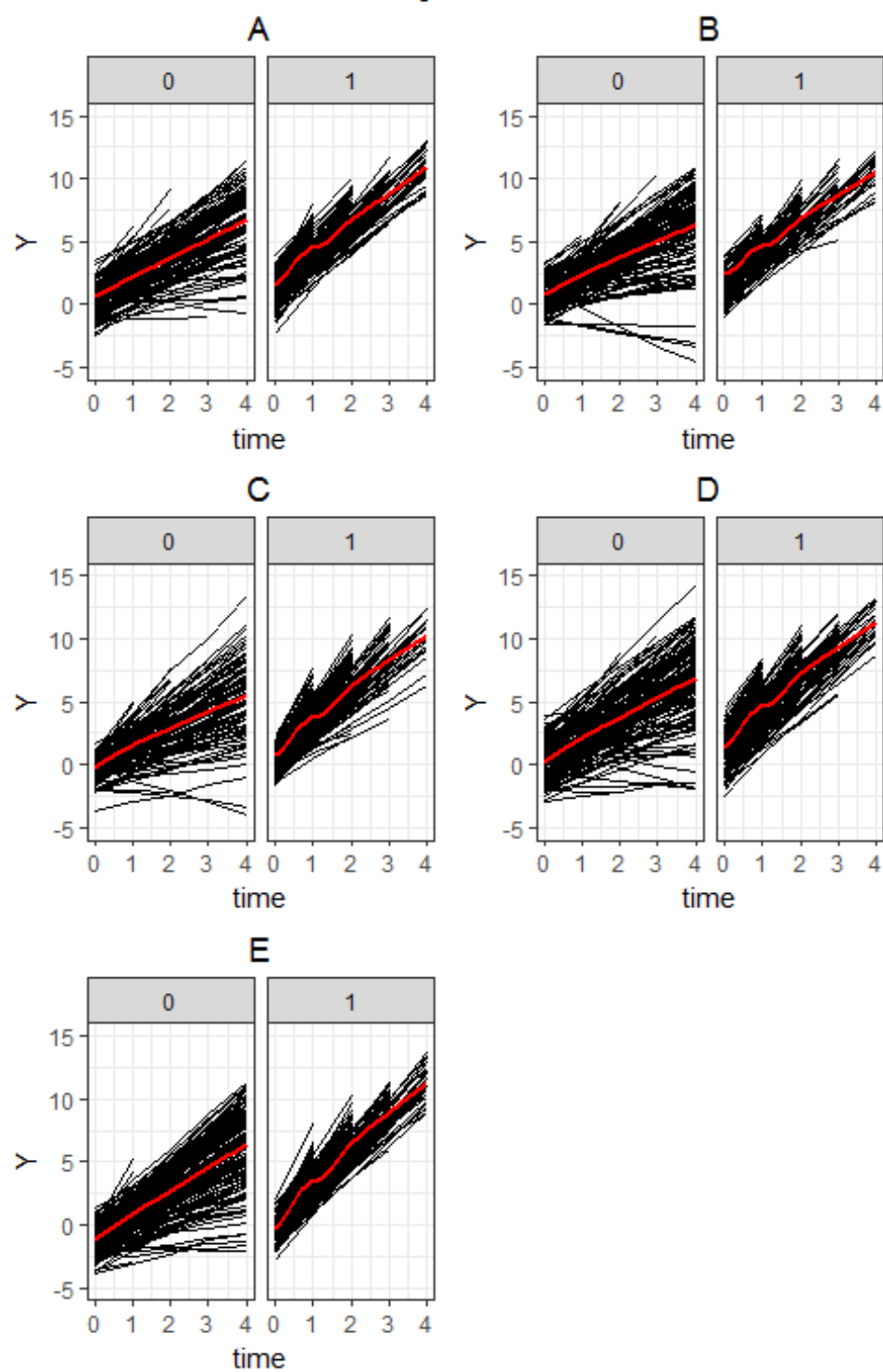
An example of using the *jointmetaplotall* function using the plots already generated from the simulated dataset *simdat* is:

```
studyplotsall<-jointmetaplotall(plotlist = studyplots, ncol = 2,  
  top = "Example Data", type = "Both")
```

To extract the grid of study specific longitudinal profiles:

```
studyplotsall$longall
```

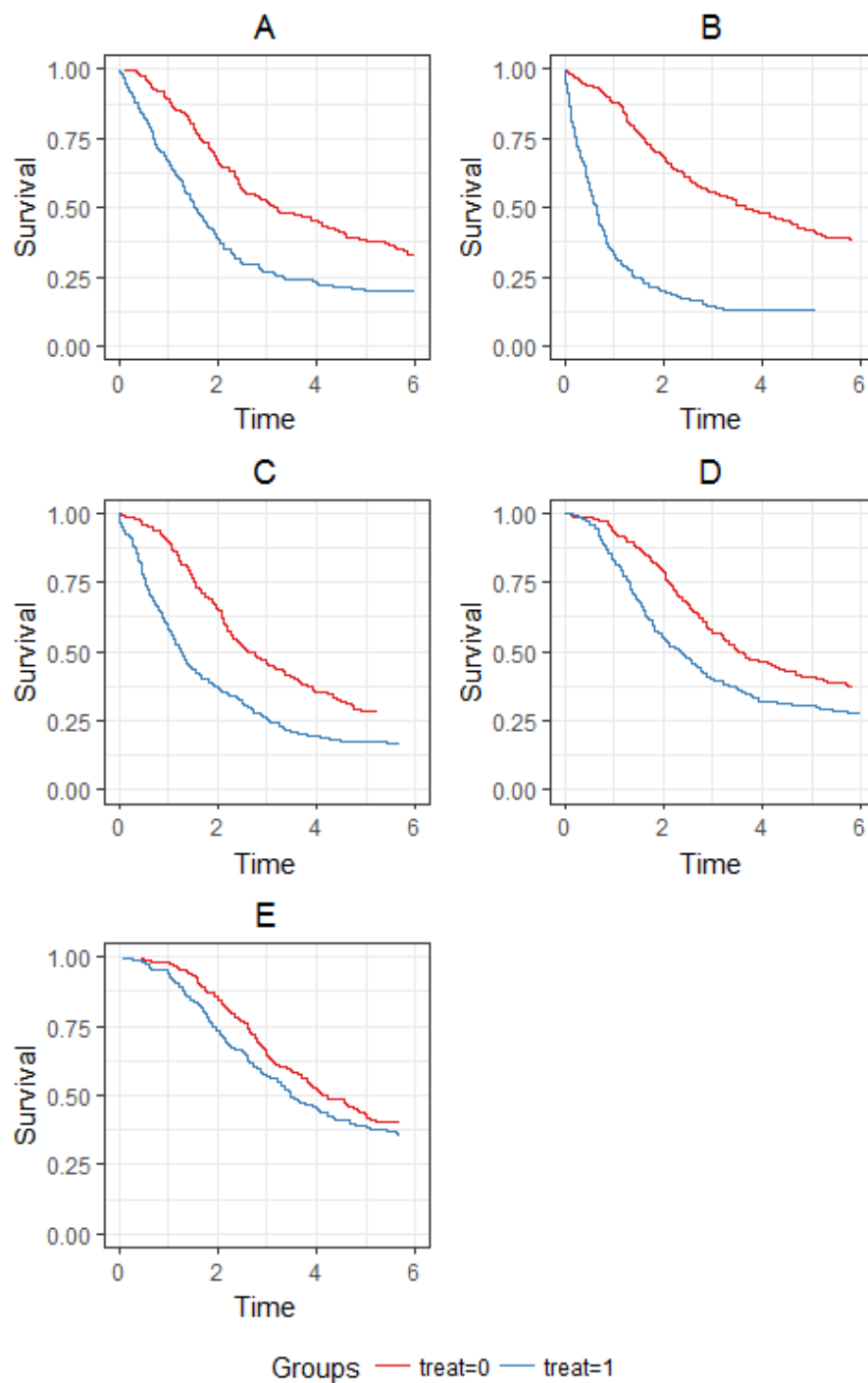
## Example Data



To extract the grid of study specific Kaplan-Meier plots:

```
studyplotsall$eventsall
```

## Example Data



## 4.5 Discussion

The *joinermeta* package contains a range of functions designed for use during the meta-analysis of multi-study joint longitudinal and time-to-event data. Functions include multi-study data simulation functions, data plotting functions, and functions to aid the preparation of data into the required format. Functions for analysis of multi-study joint data include a function for use in the second stage of a two-stage MA of joint data designed to take joint model fits from the two main joint modelling packages in R, and output meta-analyses for each specified parameter of interest. A function is also provided to allow users to fit a one-stage model to multi-study joint data, which allows users to account for variation between units at different levels of the data to be accounted for in a range of ways. As well as individual level (level 2) random effects, study level (level 3) random effects can also be included, as well as inclusion of fixed interaction terms, and the option to stratify the baseline hazard by study.

In Chapter 5, the methodology used to implement the functions in the package is described. The package is then used when performing an analysis of the INDANA dataset (Chapter 6), and also to complete simulation studies investigating both two-stage and one-stage MA methods under a variety of scenarios (Chapter 7). Planned future work to expand the package is discussed in Chapter 8.



## Chapter 5 : Methodology Supporting the Implementation of the *joinerMeta* Package

---

During this chapter the methodology used in the implementation of the R package *joinerMeta* and its functions, which were described in Chapter 4, is discussed. Details of employing the package's functions in the R programming environment are not reiterated. The chapter begins with a description of the methodology of importance to the two-stage MA and the one-stage MA functions respectively, followed by discussion of the methodology used to simulate multi-study joint data.

### 5.1 Analysis Functions

This section will cover only the mechanics of how the meta-analytic joint models are fitted. For a discussion of proposed methods for the one or two-stage MA of joint data, see Chapter 3.

#### 5.1.1 Methods for Two-stage Meta-Analyses

The function *jointmeta2* (introduced in Section 4.3.1) takes single study joint model fits from packages *joiner* and *JM*, extracts the specified parameters of interest and outputs the results of fixed and random effects meta-analyses. Standard meta-analytic techniques are employed; inverse variance weightings are used when pooling parameters, with the DerSimonian and Laird approach for the random effects MA [10].

In Section 4.3.1 it was noted that pooling joint model fits from the *joiner* package gave output containing only *survMA.direct*, whereas pooling fits from the *JM* results in output containing both *survMA.direct* and *survMA.overall*. This is due to difference in association structures available in each package. Consider a joint model of specification in equation (17):

$$\begin{aligned} Y_{ki} &= X_{ki}\beta_{1k} + Z_{ki}^{(2)}b_{ki}^{(2)} + \epsilon_{ki} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(X_{ki}\beta_{2k} + W_{2ki}(t)) \\ W_{2ki}(t) &\propto W_{1ki}(t) = \alpha_k^{(2)} \left( Z_{ki}^{(2)}b_{ki}^{(2)} \right) \end{aligned} \tag{17}$$

In the above joint model, the association structure involves only zero mean random effects, no fixed effects. As such, the effect of covariates of interest on the risk of event is fully quantified

(directly estimated) by the coefficient of the covariate when included as a time-to-event sub-model fixed effect, held in vector  $\beta_2$ . The *joiner* package fits joint models of this type, with association structures that involve only zero mean random effects.

Alternatively consider a joint model of formulation:

$$\begin{aligned} Y_{ki} &= X_{ki}\beta_{1k} + Z_{ki}^{(2)}b_{ki}^{(2)} + \varepsilon_{ki} = m_{ki} + \varepsilon_{ki} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(X_{ki}\beta_{2k} + W_{2ki}(t)) \\ W_{2ki}(t) &\propto W_{1ki}(t) = \alpha_k^{(2)} f(m_{ki}) \end{aligned} \quad (18)$$

Equation (18) contains a joint model whose sub-models are linked through an association structure that involves some function of the fixed and random effects of the longitudinal sub-model. The *JM* package allows joint models to be fitted that link sub-models using both the fixed and random effects. In cases such as this, if a covariate is present both in the longitudinal and the time-to-event sub-model, the overall effect of the covariate on the risk of an event is made up of both a direct and indirect effect [87]. As an example, consider the following model:

$$\begin{aligned} Y_{ki} &= \beta_{10k} + \beta_{11k} \text{time}_{ki} + \beta_{12k} \text{treat}_{ki} + b_{ki0}^{(2)} + b_{ki1}^{(2)} \text{time}_{ki} + \varepsilon_{ki} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{21k} \text{treat}_{ki} + \beta_{22k} \text{gender}_{ki} + W_{2ki}(t)) \\ W_{2ki}(t) &= \alpha_{k\_current} (\beta_{10k} + \beta_{11k} \text{time}_{ki} + \beta_{12k} \text{treat}_{ki} + b_{ki0}^{(2)} + b_{ki1}^{(2)} \text{time}_{ki}) \end{aligned} \quad (19)$$

If the effect of treatment assignment  $\text{treat}_{ki}$  on the risk of an event is considered, a direct effect results from the covariate being included in the time-to-event sub-model ( $\beta_{21k}$ ), but an indirect effect is also present through the association structure ( $\alpha_{k\_current}\beta_{12k}$ ). The overall effect of treatment on this risk of the event is then the sum of the direct and indirect effects ( $\beta_{21k} + \alpha_{k\_current}\beta_{12k}$ ). If a covariate is included in the time-to-event sub-model but not in the longitudinal sub-model, regardless of the fact that the association structure involves both fixed and random effects, the overall effect will equal the direct effect. An example of this is the inclusion of gender as a covariate in the time-to-event sub-model in equation (19). As it is not present in the longitudinal sub-model, the direct effect ( $\beta_{22k}$ ) equals the overall effect.

The case where a *JM* fit uses an association structure that involves both the current value and current slope of the longitudinal trajectory is more complex:

$$\begin{aligned}
Y_{ki} &= \beta_{10k} + \beta_{11k} \mathbf{time}_{ki} + \beta_{12k} \mathbf{treat}_{ki} + \beta_{13k} \mathbf{treat}_{ki} * \mathbf{time}_{ki} + b_{ki0}^{(2)} \\
&\quad + b_{ki1}^{(2)} \mathbf{time}_{ki} + \varepsilon_{ki} \\
Y_{ki} &= \mathbf{m}_{ki} + \varepsilon_{ki} \\
\lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{21k} \mathbf{treat}_{ki} + W_{2ki}(t)) \\
W_{2ki}(t) &= \alpha_{k\_current} f(\mathbf{m}_{ki}) + \alpha_{k\_slope} f'(\mathbf{m}_{ki})
\end{aligned} \tag{20}$$

In the above model, (equation (20)), the direct effect of treatment on the risk of an event quantified by  $\beta_{21k}$ . Indirect effects of treatment on risk of an event are present through the association structure equaling  $\alpha_{k\_current}\beta_{12k} + \alpha_{k\_slope}\beta_{13k}$ , giving an overall effect of treatment on risk of an event of  $\beta_{21k} + \alpha_{k\_current}\beta_{12k} + \alpha_{k\_slope}\beta_{13k}$ . As detailed in Chapter 4, the *jointmeta2* function can handle such as case of mixed current value and slope association structure.

As well as estimates of the effects of covariates, to perform a MA, estimates of the variability of the effects are required. For cases where the overall effect consists of both a direct and indirect effect this can be estimated using the delta method [201], which approximates the standard error of a function of a random variable by expanding the function around its mean and (by using a first order Taylor approximation) estimates the variance of the function. An example of this calculation for joint models using the Stata software [84], using the *nLcom* function is given by Crowther et al [160]. The same procedure can be implemented in R through the *msm* package, using the *delTamethod* function. This is automatically completed in the *jointmeta2* function.

### 5.1.2 Methods for One-Stage Meta-Analyses

This section discusses the methodology used by the one-stage multi-study joint modeling function *jointmeta1* available in the *joinermeta* package and focuses on the mechanics behind the model fitting function.

#### 5.1.2.1 General Likelihood formulation

Rizopoulos [52] notes that due to the presence of random effects in the joint model, unlike the Cox model, estimation cannot be based solely on the partial likelihood, and must instead use the full likelihood. In this section, the likelihood stated for the one study case in Wulfsohn and Tsiatis [78] is extended to a multi-study case.

The nesting of longitudinal measurements within individuals, and individuals within studies must be considered. Extending the single study case, random effects could be included at either the study or the individual level. The complete data is defined to be  $\Omega$ , where  $\Omega = (T_{Ski}, \Delta_{ki}, Y_{ki}, t_{ki}, \mathbf{b}_k^{(3)}, \mathbf{b}_{ki}^{(2)})$ , with  $T_{Ski}$  representing the survival time for individual  $i$  in study  $k$ ,  $\Delta_{ki}$  is the censoring indicator,  $Y_{ki}$  and  $t_{ki}$  are the vectors of longitudinal measurements and time points, while  $\mathbf{b}_k^{(3)}$  and  $\mathbf{b}_{ki}^{(2)}$  are the study and individual level random effects. Equation (21) shows the complete or full likelihood (referred to as  $L(\Omega)$ ) of the joint model for multiple studies.

In equation (21), the notation introduced in Chapter 3 is employed.  $K$  is the total number of included studies (with study indicator  $k = 1 \dots K$ ). The total number of individuals in study  $k$  is denoted by  $n_k$ . The total number of longitudinal measurements recorded for individual  $i$  from study  $k$  is denoted by  $m_{ki}$ . Definitions for each of the functions that contribute to  $L(\Omega)$  follow:

$$\prod_{k=1}^K \left( \int_{-\infty}^{\infty} f(\mathbf{b}_k^{(3)} | \mathbf{A}) \prod_{i=1}^{n_k} \left( \int_{-\infty}^{\infty} f(Y_{kij} | \boldsymbol{\beta}_1, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(2)}, \boldsymbol{\sigma}_e^2) \right) f(\mathbf{b}_{ki}^{(2)} | \mathbf{D}) f(T_{Ski}, \Delta_{ki} | \boldsymbol{\beta}_2, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(2)}, \boldsymbol{\alpha}) d\mathbf{b}_{ki}^{(2)} \right) d\mathbf{b}_k^{(3)} \quad (21)$$

$$f(Y_{kij} | \boldsymbol{\beta}_1, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \boldsymbol{\sigma}_e^2) = (2\pi\sigma_e^2)^{-1/2} \exp(-(Y_{kij} - \eta_{kij})^2 / 2\sigma_e^2) \quad (22)$$

$$f(T_{Ski}, \Delta_{ki} | \boldsymbol{\beta}_2, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \lambda_0, \boldsymbol{\alpha}) = \left[ \lambda_0 (T_{Ski}) \exp \left( (X_{2ki} \boldsymbol{\beta}_2) + \alpha^{(2)} (Z_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) + \alpha^{(3)} (Z_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right]^{\Delta_{ki}} \exp \left[ - \int_0^{T_{Ski}} \lambda_0(u) \exp \left( (X_{2ki} \boldsymbol{\beta}_2) + \alpha^{(2)} (Z_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) + \alpha^{(3)} (Z_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) du \right] \quad (23)$$

$$f(\mathbf{b}_k^{(3)} | \mathbf{A}) = (2\pi)^{-r/2} |\mathbf{A}|^{-1/2} \exp \left\{ - \frac{(\mathbf{b}_k^{(3)})^T \mathbf{A}^{-1} (\mathbf{b}_k^{(3)})}{2} \right\} \quad (24)$$

$$f(\mathbf{b}_{ki}^{(2)} | \mathbf{D}) = (2\pi)^{-q/2} |\mathbf{D}|^{-1/2} \exp \left\{ - \frac{(\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} (\mathbf{b}_{ki}^{(2)})}{2} \right\} \quad (25)$$

#### 5.1.2.1.1 Longitudinal sub-model component

The probability distribution function of each longitudinal measurement  $Y_{kij}$  (recorded for individual  $i$  in study  $k$  at their  $j$ th recorded time-point  $t_{kij}$ ) is given by

$f(Y_{kij} | \boldsymbol{\beta}_1, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \sigma_e^2)$ , see equation (22) for the full specification. The total contribution of the longitudinal portion to the complete likelihood for a particular individual is found by the product of this function across all time points recorded for the individual in question, i.e.  $\prod_{j=1}^{m_{ki}} f(Y_{kij} | \boldsymbol{\beta}_1, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \sigma_e^2)$ . Here,  $\boldsymbol{\beta}_1$  are the longitudinal population fixed effects,  $\mathbf{b}_{ki}^{(2)}$  are the individual specific (level 2) random effects,  $\mathbf{b}_k^{(3)}$  are the study specific (level 3) random effects (if present), and  $\sigma_e^2$  is the variance of the measurement errors (represented by  $\varepsilon_{kij}$ ). These longitudinal measurements can occur at times unique to each individual. This vector of measurement times for each individual is denoted by  $\mathbf{t}_{ki}$  and is of length  $m_{ki}$ . It is assumed that the longitudinal measure recorded at each time point for an individual can be considered normally distributed. The mean of this normal distribution is the sum of the fixed ( $\mathbf{x}_{1kij}\boldsymbol{\beta}_1$ ) and random components ( $\mathbf{z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)}$  and  $\mathbf{z}_{kij}^{(3)}\mathbf{b}_k^{(3)}$ ) of the longitudinal model (represented by  $\eta_{kij}$  where  $\eta_{kij} = \mathbf{x}_{1kij}\boldsymbol{\beta}_1 + \mathbf{z}_{kij}^{(2)}\mathbf{b}_{ki}^{(2)} + \mathbf{z}_{kij}^{(3)}\mathbf{b}_k^{(3)}$ ), with variance equal to  $\sigma_e^2$ . Note, that in the Maximisation or M-step of the Expectation Maximisation (EM) algorithm (the procedure employed in the *jointmeta1* function to fit the one-stage joint model) the random effects contribute to the mean rather than the variance of the longitudinal measurements. This is because in the Expectation or E-step, estimates of functions of the random effects are produced. In the M-step, the likelihood of the model is maximized, with functions of the random effects held constant at the estimates produced in the E-step (resulting in their contribution to the mean rather than the variance of the longitudinal measurement distribution). The use of the EM algorithm in the one-stage joint modelling function is discussed in Section 5.1.2.2.

In  $\eta_{kij}$ ,  $\mathbf{x}_{1kij}$  represents the covariates for the fixed effects at time point  $j$  for individual  $i$  in study  $k$ . The covariates over all time points for the individual are held in the design matrix  $\mathbf{X}_{1ki}$ , which will have  $m_{ki}$  rows and  $p_1$  columns (where  $p_1$  is the number of fixed effects in the longitudinal sub-model).

The covariates for the individual specific (level 2) random effects at time point  $j$  for individual  $i$  in study  $k$  are represented by  $\mathbf{z}_{kij}^{(2)}$ , and those for the study level random effects by  $\mathbf{z}_{kij}^{(3)}$ . The covariates over all time points are held in design matrices  $\mathbf{Z}_{ki}^{(2)}$  and  $\mathbf{Z}_{ki}^{(3)}$ .

respectively. The random effects themselves are represented by  $\mathbf{b}_{ki}^{(2)}$  and  $\mathbf{b}_k^{(3)}$  for the individual and study level random effects respectively. If study level random effects are not included in the model, terms involving  $\mathbf{b}_k^{(3)}$  are not present. The model assumes that the covariates assigned random effects have also been assigned fixed effects (i.e. that the columns forming design matrices  $\mathbf{Z}_{ki}^{(2)}$  and  $\mathbf{Z}_{ki}^{(3)}$  are subsets of the columns of design matrix  $\mathbf{X}_{1ki}$ ).

Throughout, only time varying covariates that can be expressed as functions of time (e.g. interactions between time and stationary variables, such as treatment group) can be assigned random effects in the longitudinal sub-model. Other time varying covariates that cannot be stated as some function of time and a stationary covariate are not allowed to be assigned random effects (e.g. weight measured at successive time points). This is due to the necessity to know the value of the random effects at  $T_{Ski}$  (the survival time of individual  $i$  in study  $k$ ); currently the value in the association structure for a time-varying covariate not calculable from a time variable and a stationary covariate cannot be approximated by the package.

#### 5.1.2.1.2 Time-to-event sub-model component

The probability distribution function of the time-to-event component is represented by  $f(T_{Ski}, \Delta_{ki} | \boldsymbol{\beta}_2, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \lambda_0, \boldsymbol{\alpha})$ , as specified in equation (23). The exact structure of this component depends on the association structure of the joint model. Only the random effects proportional association structure is currently considered, but extensions to available association structures are discussed in Section 8.3.2.2.

Throughout,  $\mathbf{X}_2$  represents the design matrix for the fixed effects in the time-to-event sub-model, and  $\boldsymbol{\beta}_2$  the fixed effect coefficients, while  $\lambda_0$  represents the unspecified baseline hazard function. The design matrix  $\mathbf{X}_2$  will have number of rows equal to the number of individuals in the analysis (total number of individuals across studies for non-stratified models ( $\sum_{k=1}^K n_k$ ), or number of individuals within a study ( $n_k$ ) for models with a stratified baseline, in which case the fixed effect design matrix would be denoted  $\mathbf{X}_{2k}$ ), and will always have  $p_2$  columns. Here  $p_2$  is the number of fixed effects in the time-to-event sub-model. It is assumed that  $\mathbf{X}_2$  contains no time variable covariates.

The survival time ( $T_{Ski}$ ) is the minimum of the true event time ( $T_{Eki}$ ) and the censoring time ( $T_{Cki}$ ) for individual  $i$  in study  $k$ . The event indicator ( $\Delta_{ki}$ ) takes a value of 1 if the individual experienced an event at  $T_{Ski}$ , and 0 otherwise. Terms shared between the sub-

models (preceded by  $\alpha$  terms) have the same definitions as in the longitudinal sub-model section. Note that any point where time is used in the shared terms in

$f(T_{Ski}, \Delta_{ki} | \beta_2, \mathbf{b}_{ki}^{(2)}, \mathbf{b}_k^{(3)}, \lambda_0, \alpha), T_{Ski}$  (or times from 0 to  $T_{Ski}$ ) are used in place of longitudinal times  $t_{kij}$ . Similarly any time varying covariates in the shared terms take values at times relating to the survival data rather than the longitudinal data.

The association terms (represented by  $\alpha$  terms) have bracketed superscripts to identify the data level they relate to. Specifically,  $\alpha^{(2)}$  denotes the association parameter for shared zero mean individual level (level 2) random effects, and  $\alpha^{(3)}$  represents the association parameter for shared zero mean study level (level 3) random effects. This function can currently only fit models that share zero mean random effects between sub-models, with common association parameter across random effects at the same level, termed random effects only proportional association.

As before, if no study level random effects are included in the model the  $\alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)})$  component is not present.

#### 5.1.2.1.3 Study specific random effects component

The probability distribution of the zero mean study specific (level 3) random effects,  $f(\mathbf{b}_k^{(3)} | \mathbf{A})$ , is given in equation (24). If no study level random effects are included in the model, then  $f(\mathbf{b}_k^{(3)} | \mathbf{A})$  is not present in equation (21) (and consequently  $\mathbf{b}_k^{(3)}$  does not require integrating out). The random effects are considered to follow a zero mean multivariate normal distribution with covariance matrix  $\mathbf{A}$  of dimension  $r$  (the number of study level random effects).

If only one study specific random effect is considered, giving  $\mathbf{A} = \sigma_A^2$  and  $r = 1$ , then equation (24) becomes a univariate distribution:

$$f(\mathbf{b}_k^{(3)} | \sigma_A^2) = (2\pi\sigma_A^2)^{-1/2} \exp\left\{-\left(\mathbf{b}_k^{(3)}\right)^2 / 2\sigma_A^2\right\}$$

#### 5.1.2.1.4 Individual level random effects component

The probability distribution function of the zero mean individual specific (level 2) random effects is given by  $f(\mathbf{b}_{ki}^{(2)} | \mathbf{D})$ , see equation (25), where  $q$  represents the number of individual specific (level 2) random effects. As with the study level random effects, the individual level random effects are assumed to follow a multivariate normal distribution with covariance matrix  $\mathbf{D}$  if  $q > 1$ :



If only one individual level random effect is included in the model, then  $\mathbf{D} = \sigma_D^2$  and  $q = 1$ , then equation (25) becomes a univariate distribution:

$$f(\mathbf{b}_{ki}^{(2)} | \sigma_D^2) = (2\pi\sigma_D^2)^{-1/2} \exp\left\{-\left(\mathbf{b}_{ki}^{(2)}\right)^2 / 2\sigma_D^2\right\}$$

The *jointmeta1()* function assumes that all models fitted will have at least one individual level random effect, and so  $f(\mathbf{b}_{ki}^{(2)} | \mathbf{D})$  will always be present in the likelihood.

#### 5.1.2.2 EM algorithm

For the single study case, the joint model is commonly fitted using the EM algorithm [197]; this method is also employed in this thesis to fit multi-study joint models. Discussion of the use of the EM algorithm with joint models for longitudinal and time-to-event data is given in Wulfsohn and Tsiatis [78] and Rizopoulos [52]. The EM algorithm fits a model by iterating between two steps – the Expectation or E-step and the Maximisation or M-step. In the E-step, estimates of the value of various functions of random effects included in the model are calculated. These estimates are then used in the M-step to determine the estimates of other parameters, for example covariance matrices and baseline hazard functions.

The EM algorithm is simple to implement and applicable to a wide range of problems [197]. It has already been shown to work for the case of joint models [52, 78]. Here, the EM methods proposed for the single study in Wulfsohn and Tsiatis [78] are extended to the case with multilevel data and random effects at more than one level.

Despite the EM algorithm being a stable, widely applicable and reliable method, issues have been identified with the use of the EM algorithm. One general issue is that whilst the EM algorithm is applicable to a wide range of situations, it can be slow to converge [202, 203]. Slow convergence is a particular issue here due to the large sizes of multi-study datasets. In addition, in order to work well the EM algorithm benefits from a good choice of starting values. These are easily obtained in the joint modelling case from separate longitudinal and time-to-event fits calculated before the main joint model fit, using the *Lme4* [204] and *survival* [195] packages respectively.

Some alternatives to the EM algorithm in the joint modelling context exist. For example Crowther et al 2012 [159] fit fully parametric joint models, allowing the use of Newton-Raphson methods to directly maximise model parameters. Alternatively, Bayesian methods such as MCMC can be used to fit joint models, such methods are implemented in

the R package *JMbayes* [82], or EM algorithms with an MC step, as employed in the *joinerML* package [181]. Potential improvements to model fitting are discussed as extension work in Chapter 8.

#### 5.1.2.2.1 Expectation Step

As mentioned, in the expectation step or E-step, expectations of functions of the random effects included in the one-stage joint model are calculated. These are then used in the M-Step to calculate the values of model parameters.

Gaussian quadrature methods approximate integrals of functions with respect to a given distributions kernel, using a weighted average of the integral at certain abscissa or locations (denoted by  $a$ , see Stroud and Secrest [205]). Weights for these abscissa, denoted by  $\pi$  terms, can be obtained from existing tables (e.g. Abromowitz and Stegun [206]) or from algorithms (see Golub [207] and Golub and Welsch [208]).

Gauss-Hermite quadrature is used is to estimate the functions of random effects needed for model fitting. This method uses a weighting function of  $e^{-x^2}$  (see Abromowitz and Stegun [206]). The major joint modelling paper Wulfsohn and Tsiatis [78] used Gauss-Hermite quadrature to evaluate functions of random effects in their joint model.

An issue with Gauss-Hermite quadrature is that if the distribution being approximated has a peak that lies far from zero, or if the spread of the distribution is different to the kernel being used in the quadrature, the quadrature may not give good results unless a large number of quadrature points are used [209].

Adaptive Gauss-Hermite quadrature provides a solution to the issue of the distribution being approximated having a different spread and location to the kernel being used in the quadrature (see Liu and Pierce [210], and Pinheiro and Bates [209]). In adaptive Gauss-Hermite quadrature the locations and spread of the abscissa are adjusted at each iteration of the model fitting process so that they more closely reflect the shape of the integral of interest. This method has the benefit of requiring fewer quadrature points, because they are placed where they are most needed (see Pinheiro and Bates [209] for simulations). However, as the points are rescaled at each iteration, this method comes with a higher computational burden.

An example of use of adaptive Gauss-Hermite quadrature in a joint modelling context is given by Crowther et al [159] for flexible parametric joint models. By using parametric models they were able to estimate model parameters using Newton-Raphson rather than

EM algorithm methods; however adaptive quadrature was used to estimate random effects. They compared results from non-adaptive Gauss-Hermite quadrature with 5 and 15 quadrature points to an adaptive method with 5 quadrature points. The adaptive approach was established as superior over the non-adaptive due to the reduced number of quadrature points required but the need to evaluate model estimates using different numbers of quadrature points to ensure stability was highlighted, a point reiterated by Lesaffre and Spiessens [211].

Current adaptive Gauss-Hermite quadrature methods for joint longitudinal and time-to-event models have been applied to data with two levels (longitudinal measurements nested within individuals). However, the data considered here has another level of nesting, namely that longitudinal measurements are nested within individuals, who are then nested within studies. Therefore Gauss-Hermite quadrature data with more than two levels must be examined.

A useful paper that describes the process and equations of using adaptive Gaussian quadrature for multilevel models is Rabe-Hesketh et al [212] who describe in detail the maximum likelihood estimation of limited and discrete dependent variable models which contain nested random intercepts and covariates using adaptive Gauss-Hermite quadrature. Other useful papers include Pinheiro and Chao [213], who demonstrate the expansion of adaptive Gaussian quadrature for generalized linear models [209] to cases with data nested in multiple levels. They note that by exploiting the structure of the integrand in a generalised linear mixed effect model, they can apply a transformation so that the quadrature becomes multiple applications of one dimensional quadrature rules. Methods for multilevel generalized linear models are also discussed by Rabe-Hesketh [214], where adaptive quadrature is again discussed for generalized linear models with more than two levels and both random intercepts and random coefficients. Gibbons and Hedeker [215] discuss approaches for three level probit and logistic regression models, and further discuss orthogonalisation of model parameters to aid estimation. Multilevel ordinal regression models are described by Hedeker and Gibbons [216], also by Raman and Hedeker [217]. The latter describes the modelling of three level ordinal data, with commonly distributed (i.e. same distribution across units at the same level) random effects at level 2 and level 3. These random effects were estimated using Gauss-Hermite quadrature, estimating random effects at lower levels followed by random effects at higher levels. Gauss-Hermite quadrature was also used by Crowther et al [58] for the modelling of

multilevel survival models with random effects in a meta-analytic setting. They extended the methods of Liu and Huang [202] from one to many random effects in a survival model.

These papers, whilst they do not implement joint longitudinal and time-to-event models, demonstrate the methods and required calculations for a range of models that deal with data with more than two levels, generally for cases just containing a random intercept at each level. Using these papers, an attempt was made to implement fully adaptive Gauss-Hermite quadrature for the estimation of random effects allowing for up to three random effects per level. The extension to data with three levels involves methods described in Rabe-Hesketh et al [212]. This paper highlighted that adaptive quadrature cannot be directly applied to a model with random effects for three level data, because the posterior density of a random effect being estimated is conditional on the random effects not yet evaluated as well as random effects at higher levels. They solve this issue through transformation of the random effects. Attempts were made to implement fully adaptive quadrature in the `joineRmeta` package. However, the methods presented in Rabe-Hesketh et al [212] proved difficult to follow, as notation changed several times during the paper, and the link between different steps of the transformations were unclear. As such, implementation of this method with joint models resulted in many model fitting issues that were difficult to resolve, and convergence of joint models could not be consistently achieved. As such, completion of implementation of fully adaptive quadrature in the package has been postponed for future research (see Section 8.3.2.2)

In a paper in 2012 [196], Rizopoulos suggested pseudo-adaptive Gauss-Hermite Quadrature as an alternative to fully adaptive quadrature for the fast fitting of joint models. In this variation, the abscissa were relocated and rescaled only once at the start of the joint model fitting process rather than at each iteration. The paper proposed fitting a standalone longitudinal model (already required to find starting points for the EM algorithm when fitting a joint model), and used empirical Bayes estimates of the locations of random effects of this longitudinal model (the conditional modes of the random effects) to relocate and rescale the abscissa to estimate the random effects in the full joint model. The paper tested the pseudo adaptive method against the fully adaptive method, and noted that the pseudo adaptive method gave similar results but was much faster than the fully adaptive. This pseudo adaptive method is currently implemented as the default quadrature option in the R package *JM* [81].

Whilst issues were encountered implementing fully adaptive quadrature for three level joint data, the pseudo adaptive procedure was easily extended. As such, the *joinerMeta* package currently relies on the pseudo adaptive Gauss-Hermite quadrature methods presented in Rizopoulos [196], using estimates of the conditional modes of the random effects available from the initial longitudinal model fit necessary to obtain suitable starting values for the EM algorithm. The procedure provides a fast and accurate method of estimation without a large computational burden, both important considerations for a one-stage meta-analysis with potentially large datasets.

Wulfsohn and Tsiatis [78] described in their paper that during the E-step of the EM algorithm it was necessary to find the expected value of the random effects and various functions of the random effects. They showed that for the model presented in their paper, with only two levels of nesting in the model (longitudinal measurements nested within individuals), the expectation of some function  $h$  of the random effects, can be expressed by equation (26):

$$\mathbb{E}[h(\mathbf{b}_i)|T_{Si}, \Delta_i, \mathbf{Y}_i, \mathbf{t}_i] = \frac{\int_{-\infty}^{\infty} h(\mathbf{b}_i) f(T_{Si}, \Delta_i | \hat{\boldsymbol{\beta}}_2, \mathbf{b}_i, \hat{\lambda}_0, \hat{\boldsymbol{\alpha}}) f(\mathbf{b}_i | \mathbf{Y}_i, \mathbf{t}_i, \hat{\boldsymbol{\beta}}_1, \hat{D}, \hat{\sigma}_e^2)}{\int_{-\infty}^{\infty} f(T_{Si}, \Delta_i | \hat{\boldsymbol{\beta}}_2, \mathbf{b}_i, \hat{\lambda}_0, \hat{\boldsymbol{\alpha}}) f(\mathbf{b}_i | \mathbf{Y}_i, \mathbf{t}_i, \hat{\boldsymbol{\beta}}_1, \hat{D}, \hat{\sigma}_e^2)} \quad (26)$$

As some of the models considered here contain random effects across several levels (individual level and study level), formulations of the above equation must be considered for cases where the model contains both  $\mathbf{b}_{ki}^{(2)}$  terms and  $\mathbf{b}_k^{(3)}$  terms, or only  $\mathbf{b}_{ki}^{(2)}$  terms.

As the EM algorithm fits the model using an iterative process, when using expressions such as that shown in equation (26) to estimate functions of the study level random effects, the individual level random effects are held constant at their estimated value from the last iteration (or their value estimated from the standalone longitudinal model in the first iteration). Similarly when the individual level random effects are estimated in the E-step, study level random effects (if included), are held at their estimated value from the last iteration. In the future, implementation of adaptive Gaussian quadrature will allow random effects at any level to be estimated simultaneously (see Section 8.3.2.2).

#### 5.1.2.2.1.1 Estimation of functions of study level random effects

If the model specified to be fitted contains study level random effects, functions of the study level random effects must be estimated during the E-step. When estimating such functions, all data relating to each study level (level 3) unit is considered, which contains

data from multiple individuals (level 2 units). As mentioned, individual level random effects are held constant at their current estimate, denoted  $\widehat{\mathbf{b}}_{ki}^{(2)}$ , when estimating study level random effects.

In order to estimate the function corresponding to the expression in equation (26) for the case where the model contains study level random effects, the expression

$f(\mathbf{b}_k^{(3)} | T_{Sk}, \Delta_k, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)})$  must be defined. Using Bayes' theorem that  $P(A|B) = P(A \cap B)/P(B)$ :

$$f(\mathbf{b}_k^{(3)} | T_{Sk}, \Delta_k, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) = \frac{f(\mathbf{b}_k^{(3)}, T_{Sk}, \Delta_k | Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)})}{f(T_{Sk}, \Delta_k | Y_{ki}, \mathbf{t}_{ki}, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)})}$$

Then using  $P(B) = \int P(B|A)P(A) dA$  and  $P(B \cap A) = P(B|A)P(A)$ :

$$\begin{aligned} & f(\mathbf{b}_k^{(3)} | T_{Sk}, \Delta_k, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) \\ &= \frac{f(T_{Sk}, \Delta_k | \mathbf{b}_k^{(3)}, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) f(\mathbf{b}_k^{(3)} | Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)})}{\int f(T_{Sk}, \Delta_k | \mathbf{b}_k^{(3)}, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) f(\mathbf{b}_k^{(3)} | Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) d\mathbf{b}_k^{(3)}} \end{aligned}$$

Therefore, the expression used to calculate the value of functions of the study level random effects is given by:

$$\begin{aligned} & \mathbb{E} \left[ h(\mathbf{b}_k^{(3)}) | T_{Sk}, \Delta_k, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)} \right] \\ &= \frac{\int h(\mathbf{b}_k^{(3)}) f(T_{Sk}, \Delta_k | \mathbf{b}_k^{(3)}, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) f(\mathbf{b}_k^{(3)} | Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) d\mathbf{b}_k^{(3)}}{\int f(T_{Sk}, \Delta_k | \mathbf{b}_k^{(3)}, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) f(\mathbf{b}_k^{(3)} | Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)}) d\mathbf{b}_k^{(3)}} \end{aligned} \quad (27)$$

In equation (27),  $h(\mathbf{b}_k^{(3)})$  is some function of the study level random effects to be estimated. The specification of  $f(T_{Sk}, \Delta_k | \mathbf{b}_k^{(3)}, Y_k, \mathbf{t}_k, \widehat{\Omega}, \widehat{\mathbf{b}}_{ki}^{(2)})$ , the time-to-event component of the full likelihood, is already known (see equation (23)).

If adaptive Gaussian quadrature was employed in the E-step, as in Wulfsohn and Tsiatis [78], the distribution of the study level random effects given the longitudinal measurements would be calculated at each iteration, in order to determine the values of the abscissa to employ during the quadrature. However, as pseudo-adaptive Gaussian quadrature is employed, the abscissa are calculated once before beginning the iterative

process, based on the locations of the random effects estimated by the standalone longitudinal model.

The abscissa are rescaled at the start of the iterations to fit the joint model using the following expression:

$$\mathbf{a}_{(3)k} = \hat{\mathbf{b}}_k^{(3)} + \sqrt{2}\hat{\mathbf{B}}_k^{-1}\mathbf{a} \quad (28)$$

In equation (19),  $\mathbf{a}$  denotes the vector of standard abscissa locations [205], and  $\hat{\mathbf{b}}_k^{(3)}$  are the conditional nodes of the study level random effects produced by the standalone longitudinal model fit. The matrix  $\hat{\mathbf{B}}_k$  is the Choleski decomposition of the matrix of the negative second derivative of the log-likelihood  $-\partial^2 \ell(\Omega)/\partial \mathbf{b}_k^{(3)} \partial \mathbf{b}_k^{(3)T}$ , which is the covariance matrix of the estimates of the study level random effects from the standalone longitudinal model fit. The expression produces abscissa for the study level or level 3 random effects, denoted by  $\mathbf{a}_{(3)k}$ . Once values for the abscissa for the study level random effects have been calculated, these abscissa values are input in place of any occurrence of the study level random effects. Given these abscissa values, and the definitions of the functions that constitute equation (27), estimates of any function  $h(\mathbf{b}_k^{(3)})$  of the study level random effects can be calculated, which are then used in the M-step to update estimates of the model parameters.

#### 5.1.2.2.1.2 Estimation of functions of individual level random effects

If the model to be fitted contains individual level random effects, functions of the individual level random effects must be estimated during the E-step (these are required to maximize estimates of model parameters in the M-step). When estimating these functions, all data relating to each individual level (level 2) unit is considered, and any study level random effects are held constant at their current estimate (denoted by  $\hat{\mathbf{b}}_k^{(3)}$ ). If no study level random effects are included in the model,  $\hat{\mathbf{b}}_k^{(3)}$  terms are removed from the following calculations.

In order to estimate the function corresponding to the expression in equation (26) for the case where the model contains individual level random effects, the expression

$f(\mathbf{b}_{ki}^{(2)} | \mathbf{T}_{Ski}, \Delta_{ki}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\Omega}, \hat{\mathbf{b}}_k^{(3)})$  must be defined. Using  $P(A|B) = P(A \cap B)/P(B)$ :

$$f(\mathbf{b}_{ki}^{(2)} | \mathbf{T}_{Ski}, \Delta_{ki}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\Omega}, \hat{\mathbf{b}}_k^{(3)}) = \frac{f(\mathbf{b}_{ki}^{(2)}, \mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\Omega}, \hat{\mathbf{b}}_k^{(3)})}{f(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\Omega}, \hat{\mathbf{b}}_k^{(3)})}$$

Then using  $P(B) = \int P(B|A)P(A) dA$  and  $P(B \cap A) = P(B|A)P(A)$ :

$$\begin{aligned} & f\left(\mathbf{b}_{ki}^{(2)} | \mathbf{T}_{Ski}, \Delta_{ki}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) \\ &= \frac{f\left(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{b}_{ki}^{(2)}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) f\left(\mathbf{b}_{ki}^{(2)} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right)}{\int f\left(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{b}_{ki}^{(2)}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) f\left(\mathbf{b}_{ki}^{(2)} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) d\mathbf{b}_{ki}^{(2)}} \end{aligned}$$

Therefore, the formula needed to estimate functions of the individual level random effects is given by:

$$\begin{aligned} & \mathbb{E}\left[h\left(\mathbf{b}_{ki}^{(2)}\right) | \mathbf{T}_{Ski}, \Delta_{ki}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right] \\ &= \frac{\int h\left(\mathbf{b}_{ki}^{(2)}\right) f\left(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{b}_{ki}^{(2)}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) f\left(\mathbf{b}_{ki}^{(2)} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) d\mathbf{b}_{ki}^{(2)}}{\int f\left(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{b}_{ki}^{(2)}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) f\left(\mathbf{b}_{ki}^{(2)} | \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right) d\mathbf{b}_{ki}^{(2)}} \end{aligned} \quad (29)$$

In equation (29),  $h\left(\mathbf{b}_{ki}^{(2)}\right)$  is the function of the individual level random effects to be estimated. The specification of  $f\left(\mathbf{T}_{Ski}, \Delta_{ki} | \mathbf{b}_{ki}^{(2)}, \mathbf{Y}_{ki}, \mathbf{t}_{ki}, \hat{\boldsymbol{\Omega}}, \hat{\mathbf{b}}_k^{(3)}\right)$ , the time-to-event component of the full likelihood, is already known (see equation (23)).

As before, as pseudo-adaptive Gaussian quadrature is employed, the abscissa at which to assess the integral are calculated once at the start of the iterative process. The abscissa follow equation (30):

$$\mathbf{a}_{(2)ki} = \hat{\mathbf{b}}_{ki}^{(2)} + \sqrt{2}\hat{\mathbf{B}}_{ki}^{-1}\mathbf{a} \quad (30)$$

Here,  $\mathbf{a}$  denotes the vector of standard abscissa locations [205], and  $\hat{\mathbf{b}}_{ki}^{(2)}$  are the conditional nodes of the individual level random effects produced by the standalone longitudinal model fit. The matrix  $\hat{\mathbf{B}}_{ki}$  is the Choleski decomposition of the matrix of the negative second derivative of the log-likelihood  $-\partial^2 \ell(\boldsymbol{\Omega}) / \partial \mathbf{b}_{ki}^{(2)} \partial \mathbf{b}_{ki}^{(2)T}$ , which is the covariance matrix of the estimates of the individual level random effects from the standalone longitudinal model fit. The expression produces abscissa for the individual level (level 2) random effects, denoted by  $\mathbf{a}_{(2)ki}$ . Once values for the abscissa for the individual level random effects have been calculated, these abscissa values are input in place of any occurrence of the individual level random effects. Given these abscissa values, and the definitions of the functions that constitute equation (29), estimates of any function  $h\left(\mathbf{b}_{ki}^{(2)}\right)$  of the individual level random effects can be calculated, which are then used in the M-step to update estimates of the model parameters.



#### 5.1.2.2.2 Maximization Step

During the Maximization or M-step of the EM algorithm, the estimates of functions of random effects (calculated through pseudo-adaptive Gaussian quadrature during the E-step) are used to maximize the estimates of the model parameters. Throughout this section, the expressions and procedures used to maximize these estimates are stated. Much of the work extends either standard likelihood theory [88], or theory of joint model parameter estimation already available in the literature [52, 78].

Throughout, the expressions are stated as they would appear if study level random effects were included in the model. However, if study level random effects were not present, any components containing them would not appear in the stated expressions. Additionally, any components that are expressions involving random effects are stated as expectations of the functions, e.g.  $\mathbb{E} \left[ \exp \left( \alpha^{(3)} \left( Z_h^{(3)} \mathbf{b}_h^{(3)} \right) \right) \right]$ , highlighting that these components are functions of the random effects whose expected values have been calculated in the E-step. Finally, as the complete likelihood for models fitted can be complex, where appropriate, only the portions of the likelihood that deal with the parameters being maximized are stated, e.g. for the measurement error variance  $\sigma_e^2$  only the portion of  $\mathbf{L}(\boldsymbol{\Omega})$  that deals with the longitudinal data is shown.

##### 5.1.2.2.2.1 Estimation of the unspecified baseline hazard

Wulfsohn and Tsiatis [78] note that the baseline hazard function is only considered to take weight at an event time. This is in line with the discussion that follows the main text in Cox [50]. Additionally, the closed form estimate of the baseline hazard function in Wulfsohn and Tsiatis [78] takes the form of the Breslow estimator for the baseline hazard (see Breslow, part of the discussion of Cox [50]). The estimators in the models fitted by *joinermeta* take a similar form, based on the Breslow estimator, but summed across all the included studies:

$$\hat{\lambda}_0(u_g) = \sum_{k=1}^K \sum_{i=1}^{n_k} \frac{\Delta_{ki} \mathbb{I}(T_{Ski} = u_g)}{\sum_{h \in R(u_g)} \exp(X_{2h} \boldsymbol{\beta}_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)} \left( Z_h^{(2)} \mathbf{b}_h^{(2)} \right) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)} \left( Z_h^{(3)} \mathbf{b}_h^{(3)} \right) \right) \right]} \quad (31)$$

In equation (31),  $g$  represents a particular event (where  $g = 1, \dots, G$  is used to count through the unique event times). The indicator variable  $\mathbb{I}(T_{Ski} = u_g)$  ensures that the baseline hazard function only takes weight at times when there is an event or censoring, and multiplying it by the censoring indicator  $\Delta_{ki}$  ensures weight is only taken event times. The risk set (individuals who have not yet been censored or have not yet experienced an

event) at event time  $u_g$  is denoted by  $R(u_g)$ . For stratified models the risk set is drawn separately from each included study's population, whereas for un-stratified models the risk set is drawn from all individuals in the meta-analysis. As such, for models with a stratified baseline hazard, a separate baseline hazard will be calculated for each study included in the meta-analysis. The expression in equation (31) is given for an un-stratified baseline hazard, however the expression can be simply modified to the stratified case by changing those who contribute to the risk set. For a discussion of one-stage multi-study joint models with or without a stratified baseline hazard see Section 3.3.

#### 5.1.2.2.2 Estimation of the longitudinal fixed effect coefficients

The longitudinal sub-model fixed effects ( $\beta_1$ ) were estimated using the Ordinary Least Squares (OLS) estimator. This estimator is based on the residual sum of squares (RSS), denoted by:

$$\begin{aligned}
 & (Y - X_1\beta_1 - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}])^T (Y - X_1\beta_1 - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) \quad (32) \\
 & = ((Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) - X_1\beta_1)^T ((Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) \\
 & \quad - X_1\beta_1) \\
 & = (Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}])^T (Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) \\
 & \quad - 2\beta_1^T X_1^T (Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) + \beta_1^T X_1^T X_1 \beta_1
 \end{aligned}$$

As population parameters are being estimated, common across all included studies, the estimation procedure involves all data in the dataset. The first derivative of this expression with respect to the fixed effect coefficients  $\beta_1$  gives:

$$\frac{d}{d\beta_1} = -2X_1^T (Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}]) + 2X_1^T X_1 \beta_1 \quad (33)$$

Setting this equal to zero, and rearranging, gives the ordinary least squares estimate of the longitudinal fixed effects:

$$\hat{\beta}_1 = (X^T X)^{-1} (X^T (Y - \mathbb{E}[Z^{(2)}b^{(2)}] - \mathbb{E}[Z^{(3)}b^{(3)}])) \quad (34)$$

In the future, as the package is expanded, alternative methods to estimate the longitudinal fixed effects may be necessary (e.g. when  $\beta_1$  parameters are present in the association structure). A possible alternative is the Newton Raphson method, as described by Rizopoulos [52].

#### 5.1.2.2.3 Estimation of the time-to-event fixed effect coefficients and association parameters

The fixed effect coefficients ( $\beta_2$ ) in the time-to-event sub-model and the association parameters ( $\alpha^{(2)}$  and  $\alpha^{(3)}$ ) will be estimated using a one-step Newton Raphson method (as is currently used in *joiner* [80] , and discussed by Rizopoulos [52]). To update the estimates of these parameters, the score vector  $\mathbf{S}$  and the information matrix  $\mathbf{I}$  must be calculated.

It should be highlighted that during model fitting, the association parameters are calculated along with the fixed effect parameters. Therefore the score vector  $\mathbf{S}$  will be of length  $p_2$  (the number of time-to-event sub-model fixed effect coefficients) plus the number of association parameters (one for each level of random effects included in the model). Consequently, if the model just contains individual level random effects,  $\mathbf{S}$  will be of length  $p_2 + 1$ , whereas if the model contains both individual and study level random effects  $\mathbf{S}$  will have length  $p_2 + 2$ . Note that the software requires at least one individual level random effect to be specified in the model (to ensure the presence of parameters to share between sub-models). The first  $p_2$  elements of  $\mathbf{S}$  equal the first differentiate of the expected log-likelihood with respect to each value in the vector of time-to-event fixed effect coefficients  $\beta_2$ , given the estimated values of the functions of any random effects. The remaining elements are the first differentiate with respect to each association parameter included in the model (ordered with the association parameter for the individual level random effects first, followed by that for the study level random effects if included).

The information matrix  $\mathbf{I}$  is a square matrix that contains the negative of the second order differentiates of the expected log-likelihood, giving a matrix with number of rows and columns equal to  $p_2$  plus the number of association parameters.

The estimates of the elements of  $\beta_2$  can be updated at each iteration using equation (35), where  $\nu$  is the iteration counter, such that  $\hat{\beta}_{2(\nu-1)}$  are the estimates of the time-to-event sub-model fixed effect coefficients from iteration  $\nu - 1$ . Furthermore,  $\mathbf{S}_{(\nu-1)}(\hat{\beta}_{2(\nu-1)})$  is the score function, and  $\mathbf{I}_{(\nu-1)}^{-1}(\hat{\beta}_{2(\nu-1)})$  the inverse of the information matrix, both based on the coefficient estimates  $\hat{\beta}_{2(\nu-1)}$ :

$$\hat{\beta}_{2(\nu)} = \hat{\beta}_{2(\nu-1)} + \mathbf{I}_{(\nu-1)}^{-1}(\hat{\beta}_{2(\nu-1)})\mathbf{S}_{(\nu-1)}(\hat{\beta}_{2(\nu-1)}) \quad (35)$$

The estimates of the association parameters are updated in a similar way. Below, the expressions necessary to calculate the score vector and the information matrix at each step are stated. The expressions are given for a case where both individual level and study level random effects have been included in the model (giving the score vector  $\mathbf{S}$  a length of  $p_2 + 2$ , and the information matrix  $\mathbf{I}$  row and column dimensions equal to  $p_2 + 2$ ).

The first  $p_2$  elements of  $\mathbf{S}$ , relate to the coefficients of the fixed effects in the time-to-event sub-model, and will have the form:

$$\begin{aligned} \mathbf{S}(\beta_{2g}) &= \sum_{k=1}^K \sum_{i=1}^{n_k} \Delta_{ki} x_{2kig} \\ &\quad - \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{T_{ski}} \lambda_0(u) (x_{2kig}) \exp(\mathbf{X}_{2ki} \boldsymbol{\beta}_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)} (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)} (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du \end{aligned}$$

In the above equation,  $g$  takes values 1 to  $p_2$ . The remaining elements of the score vector relate to the association parameters for the individual level random effects ( $\alpha^{(2)}$ ), and (if included) the study level random effects ( $\alpha^{(3)}$ ), and take the following forms:

$$\begin{aligned} \mathbf{S}(\alpha^{(2)}) &= \sum_{k=1}^K \sum_{i=1}^{n_k} \Delta_{ki} \mathbb{E}[(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)})] \\ &\quad - \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{T_{ski}} \lambda_0(u) \exp(\mathbf{X}_{2ki} \boldsymbol{\beta}_2) \mathbb{E}[(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \exp(\alpha^{(2)} (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}))] \mathbb{E}[\exp(\alpha^{(3)} (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}))] du \end{aligned}$$

$$\begin{aligned} \mathbf{S}(\alpha^{(3)}) &= \sum_{k=1}^K \sum_{i=1}^{n_k} \Delta_{ki} \mathbb{E}[(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)})] \\ &\quad - \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{T_{ski}} \lambda_0(u) \exp(\mathbf{X}_{2ki} \boldsymbol{\beta}_2) \mathbb{E}[\exp(\alpha^{(2)} (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}))] \mathbb{E}[(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \exp(\alpha^{(3)} (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}))] du \end{aligned}$$

The information matrix  $\mathbf{I}$  will have the following form:

$$\mathbf{I} = \begin{pmatrix} I(\beta_{21}) & \cdots & I(\beta_{21}, \beta_{2p_2}) & I(\beta_{21}, \alpha^{(2)}) & I(\beta_{21}, \alpha^{(3)}) \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ I(\beta_{21}, \beta_{2p_2}) & \cdots & I(\beta_{2p_2}) & I(\beta_{2p_2}, \alpha^{(2)}) & I(\beta_{2p_2}, \alpha^{(3)}) \\ I(\beta_{21}, \alpha^{(2)}) & \cdots & I(\beta_{2p_2}, \alpha^{(2)}) & I(\alpha^{(2)}) & I(\alpha^{(2)}, \alpha^{(3)}) \\ I(\beta_{21}, \alpha^{(3)}) & \cdots & I(\beta_{2p_2}, \alpha^{(3)}) & I(\alpha^{(2)}, \alpha^{(3)}) & I(\alpha^{(3)}) \end{pmatrix}$$

Note that any row or column containing the study level association parameter  $\alpha^{(3)}$  will not be present if study level random effects are not included in the model. The elements in the information matrix correspond to different expressions depending on their location in the matrix. The first  $p_2$  values on the diagonal have form:

$$I(\beta_{2g}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) (x_{2kig})^2 \exp(X_{2ki}\beta_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

Whilst the off-diagonals in the first  $p_2$  rows and columns of the information matrix take form:

$$I(\beta_{2g}, \beta_{2f}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) (x_{2kig} x_{2kif}) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

In the above expressions,  $g$  and  $f$  take values between 1 and  $p_2$  (with  $g \neq f$ ). The remaining on-diagonals take form:

$$I(\alpha^{(2)}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)})^2 \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

$$I(\alpha^{(3)}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)})^2 \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

The remaining off diagonal elements take the following forms, depending on their placement in the information matrix (see above for matrix structure).

$$I(\beta_{2g}, \alpha^{(2)}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) (x_{2kig}) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

$$I(\beta_{2g}, \alpha^{(3)}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) (x_{2kig}) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

$$I(\alpha^{(2)}, \alpha^{(3)}) = \sum_{k=1}^K \sum_{i=1}^{n_k} \int_0^{TSki} \lambda_0(u) \exp(X_{2ki}\beta_2) \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \exp \left( \alpha^{(2)}(\mathbf{Z}_{ki}^{(2)} \mathbf{b}_{ki}^{(2)}) \right) \right] \mathbb{E} \left[ (\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \exp \left( \alpha^{(3)}(\mathbf{Z}_{ki}^{(3)} \mathbf{b}_k^{(3)}) \right) \right] du$$

Using these expressions that make up the score vector  $\mathbf{S}$  and information matrix  $\mathbf{I}$  along with the Newton-Raphson procedure stated in equation (35), the estimates for the time-to-event sub-model fixed effect coefficients  $\boldsymbol{\beta}_2$ , and the association parameters  $\alpha^{(2)}$  and  $\alpha^{(3)}$ , can be updated at each iteration.

#### 5.1.2.2.2.4 Estimation of the longitudinal measurement error variance

The maximum likelihood estimator for the measurement error variance  $\sigma_e^2$  in the longitudinal sub-model is estimated using standard maximum likelihood theory. Displaying only the portion of likelihood that involves this parameter (with  $\boldsymbol{\eta}_{ki} = \mathbf{X}_{1ki}\boldsymbol{\beta}_1 +$

$$\mathbb{E}[\mathbf{Z}_{ki}^{(2)}\mathbf{b}_{ki}^{(2)}] + \mathbb{E}[\mathbf{Z}_{ki}^{(3)}\mathbf{b}_{ki}^{(3)}]):$$

$$\begin{aligned} L(\boldsymbol{\Omega}) &= \prod_{k=1}^K \prod_{i=1}^{n_k} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(Y_{ki} - \boldsymbol{\eta}_{ki})^2}{2\sigma_e^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_e^2}}\right)^{\sum_{k=1}^K n_k} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ki} - \boldsymbol{\eta}_{ki})^2\right) \end{aligned} \quad (36)$$

Giving a log-likelihood of:

$$\ell(\boldsymbol{\Omega}) = -\frac{\sum_{k=1}^K n_k}{2} \log(2\pi) - \frac{\sum_{k=1}^K n_k}{2} \log(\sigma_e^2) - \frac{1}{2\sigma_e^2} \sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ki} - \boldsymbol{\eta}_{ki})^2 \quad (37)$$

Taking the first derivative of equation (37) with respect to  $\sigma_e^2$  and setting equal to zero gives:

$$0 = -\frac{\sum_{k=1}^K n_k}{2\sigma_e^2} + \frac{1}{2\sigma_e^4} \sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ki} - \boldsymbol{\eta}_{ki})^2 \quad (38)$$

Which rearranges to give the maximum likelihood estimator of the parameter  $\sigma_e^2$ :

$$\sigma_e^2 = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} (Y_{ki} - \boldsymbol{\eta}_{ki})^2}{\sum_{k=1}^K n_k} \quad (39)$$

#### 5.1.2.2.2.5 Estimation of the covariance matrix for individual level random effects

The one-stage function available in *joinermeta* requires the existence of at least one individual level random effect in the specified model. The composition of the Maximum Likelihood Estimate (MLE) for the variance of these random effects depends on the number of random effects included in the model.

If multiple individual level random effects are included in the model, they follow a multivariate zero mean normal distribution. Therefore the maximum likelihood estimate for the covariance matrix of this distribution, denoted  $\hat{\mathbf{D}}$  is required. The methods to obtain this estimate rely on the methods presented in Wulfsohn and Tsiatis [78] to estimate parameters for the joint model, and Anderson and Olkin [218] who discuss methods to obtain maximum likelihood estimates for multivariate normal distributions.

In the complete data likelihood the only component to directly involve the covariance matrix  $\mathbf{D}$  for the individual specific random effects is  $f(\mathbf{b}_{ki}^{(2)} | \mathbf{D})$ , (equation (40)).

$$L(\boldsymbol{\Omega}) = \mathbb{E} \left[ \prod_{k=1}^K \prod_{i=1}^{n_k} (2\pi)^{-q/2} |\mathbf{D}|^{-1/2} \exp \left\{ -\frac{(\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} (\mathbf{b}_{ki}^{(2)})}{2} \right\} \right] \quad (40)$$

The expectation (denoted by  $\mathbb{E}$ ) is taken over the individual level random effects. The expected values of the required functions of the individual level random effects are calculated in the E-step of the EM algorithm. The log-likelihood is then:

$$\ell(\boldsymbol{\Omega}) = -\sum_{k=1}^K \frac{qn_k}{2} \log(2\pi) - \sum_{k=1}^K \frac{n_k}{2} \log |\mathbf{D}| - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} (\mathbf{b}_{ki}^{(2)}) \right] \quad (41)$$

In equation (41), consider  $\mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} (\mathbf{b}_{ki}^{(2)}) \right]$  as the trace of a 1 by 1 matrix, leading to:

$$\approx -\sum_{k=1}^K \frac{n_k}{2} \log |\mathbf{D}| - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \text{tr} \left( \mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} (\mathbf{b}_{ki}^{(2)}) \right] \right) \quad (42)$$

Using that fact that  $\text{tr}(\mathbf{GH}) = \text{tr}(\mathbf{HG})$ , rearranges to:

$$\approx -\sum_{k=1}^K \frac{n_k}{2} \log |\mathbf{D}| - \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{n_k} \text{tr} \left( \mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)}) (\mathbf{b}_{ki}^{(2)})^T \mathbf{D}^{-1} \right] \right) \quad (43)$$

Again, using  $\text{tr}(\mathbf{GH}) = \text{tr}(\mathbf{HG})$  gives:

$$\approx -\sum_{k=1}^K \frac{n_k}{2} \log |\mathbf{D}| - \frac{1}{2} \text{tr} \left( \mathbf{D}^{-1} \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)}) (\mathbf{b}_{ki}^{(2)})^T \right] \right) \quad (44)$$

Defining  $\mathbf{P} = \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ (\mathbf{b}_{ki}^{(2)}) (\mathbf{b}_{ki}^{(2)})^T \right]$ , equation (44) is restated to:

$$\approx -\left(\sum_{k=1}^K \frac{n_k}{2}\right) \log|\mathbf{D}| - \frac{1}{2} \text{tr}(\mathbf{D}^{-1}\mathbf{P}) \quad (45)$$

From Anderson and Olkin [218], an expression  $f(\mathbf{\Sigma}, \mathbf{V}) = -\log|\mathbf{\Sigma}| - \text{tr}(\mathbf{\Sigma}^{-1}\mathbf{V})$  has derivative with respect to elements of  $\mathbf{\Sigma}$  (in matrix form) of  $-\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}^{-1}\mathbf{V}\mathbf{\Sigma}^{-1}$ . Therefore, the first derivative of equation (45), set equal to zero, (taking into account that the values of  $\sum_{k=1}^K \frac{n_k}{2}$  and  $\frac{1}{2}$  remain constant when differentiating with respect to the elements of  $\mathbf{D}$ ) is:

$$\begin{aligned} -\left(\sum_{k=1}^K \frac{n_k}{2}\right) \mathbf{D}^{-1} + \frac{1}{2} \mathbf{D}^{-1} \mathbf{P} \mathbf{D}^{-1} &= 0 \\ \frac{1}{2} \mathbf{D}^{-1} \mathbf{P} \mathbf{D}^{-1} &= \left(\sum_{k=1}^K \frac{n_k}{2}\right) \mathbf{D}^{-1} \\ \mathbf{D}^{-1} \mathbf{P} \mathbf{D}^{-1} &= \left(\sum_{k=1}^K n_k\right) \mathbf{D}^{-1} \\ \mathbf{D}^{-1} \mathbf{P} &= \left(\sum_{k=1}^K n_k\right) \mathbf{D}^{-1} \\ \mathbf{P} &= \mathbf{D} \left(\sum_{k=1}^K n_k\right) \\ \hat{\mathbf{D}} &= \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ \left( \mathbf{b}_{ki}^{(2)} \right) \left( \mathbf{b}_{ki}^{(2)} \right)^T \right]}{\sum_{k=1}^K n_k} \end{aligned} \quad (46)$$

Equation (46) states the maximum likelihood estimate for the covariance matrix  $\mathbf{D}$ , for the case where two or more individual level random effects have been included in the model to be fitted by the one-stage function. If the model contained only one individual level random effect, then  $\mathbf{D}$  would be a single value, termed  $\sigma_D^2$ ). In this case, the relevant portion of the likelihood has the form:

$$L(\mathbf{\Omega}) = \mathbb{E} \left[ \prod_{k=1}^K \prod_{i=1}^{n_k} (2\pi\sigma_D^2)^{-1/2} \exp \left( -(\mathbf{b}_{ki}^{(2)})^2 / 2\sigma_D^2 \right) \right] \quad (47)$$

Then the log-likelihood:

$$\ell(\mathbf{\Omega}) = -\sum_{k=1}^K \frac{n_k}{2} \log(2\pi) - \sum_{k=1}^K \frac{n_k}{2} \log(\sigma_D^2) - \frac{1}{2\sigma_D^2} \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ \left( \mathbf{b}_{ki}^{(2)} \right)^2 \right] \quad (48)$$

Giving a first derivative with respect to  $\sigma_D^2$  of:



$$\ell'(\boldsymbol{\Omega}) = -\sum_{k=1}^K \frac{n_k}{2\sigma_D^2} + \frac{1}{2\sigma_D^4} \sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ \left( \mathbf{b}_{ki}^{(2)} \right)^2 \right] \quad (49)$$

Which, when set to zero and rearranged gives an MLE for  $\sigma_D^2$  of:

$$\sigma_D^2 = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \mathbb{E} \left[ \left( \mathbf{b}_{ki}^{(2)} \right)^2 \right]}{\sum_{k=1}^K n_k} \quad (50)$$

#### 5.1.2.2.2.6 Estimation of the covariance matrix for study level random effects

It has already been noted that when no study specific (level 3) random effects are included in the model, the  $f(\mathbf{b}_k^{(3)}|\mathbf{A})$  term is not included in the complete data likelihood. If only one study level random effect is included in the model, then MLE of the variance of the study level random effect, denoted  $\sigma_A^2$ , can be determined as follows. The likelihood takes the form:

$$L(\boldsymbol{\Omega}) = \mathbb{E} \left[ \prod_{k=1}^K (2\pi\sigma_A^2)^{-1/2} \exp \left( -(\mathbf{b}_k^{(3)})^2 / 2\sigma_A^2 \right) \right] \quad (51)$$

Giving log-likelihood:

$$\ell(\boldsymbol{\Omega}) = -\frac{K}{2} \log(2\pi) - \frac{K}{2} \log(\sigma_A^2) - \frac{1}{2\sigma_A^2} \sum_{k=1}^K \mathbb{E} \left[ (\mathbf{b}_k^{(3)})^2 \right] \quad (52)$$

Taking the first derivative of equation (52), setting equal to zero and rearranging, gives the MLE for  $\sigma_A^2$ :

$$\sigma_A^2 = \frac{\sum_{k=1}^K \mathbb{E} \left[ (\mathbf{b}_k^{(3)})^2 \right]}{K} \quad (53)$$

If more than one study level random effect is included in the model, the relevant portion of the likelihood takes the form:

$$L(\boldsymbol{\Omega}) = \mathbb{E} \left[ \prod_{k=1}^K (2\pi)^{-r/2} |\mathbf{A}|^{-1/2} \exp \left\{ -\frac{(\mathbf{b}_k^{(3)})^T \mathbf{A}^{-1} (\mathbf{b}_k^{(3)})}{2} \right\} \right] \quad (54)$$

The MLE for the covariance matrix  $\mathbf{A}$  is derived in a similar way to that for individual level random effects matrix  $\mathbf{D}$ . Rearranging to the log-likelihood gives:

$$\ell(\boldsymbol{\Omega}) = -\frac{K}{2} \log(2\pi) - \frac{K}{2} \log|\mathbf{A}| - \frac{1}{2} \sum_{k=1}^K \mathbb{E} \left[ (\mathbf{b}_k^{(3)})^T \mathbf{A}^{-1} (\mathbf{b}_k^{(3)}) \right] \quad (55)$$

$$\approx -\frac{K}{2}\log|\mathbf{A}| - \frac{1}{2}\sum_{k=1}^K \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right)^T \mathbf{A}^{-1} \left( \mathbf{b}_k^{(3)} \right) \right]$$

As before, consider  $\mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right)' \mathbf{A}^{-1} \left( \mathbf{b}_k^{(3)} \right) \right]$  to be the trace of a 1 by 1 matrix. Using this, and the fact that  $tr(\mathbf{GH}) = tr(\mathbf{HG})$ , allows:

$$\begin{aligned} &= -\frac{K}{2}\log|\mathbf{A}| - \frac{1}{2}\sum_{k=1}^K tr \left( \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right)^T \mathbf{A}^{-1} \left( \mathbf{b}_k^{(3)} \right) \right] \right) \\ &= -\frac{K}{2}\log|\mathbf{A}| - \frac{1}{2}\sum_{k=1}^K tr \left( \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right) \left( \mathbf{b}_k^{(3)} \right)^T \mathbf{A}^{-1} \right] \right) \\ &= -\frac{K}{2}\log|\mathbf{A}| - \frac{1}{2}tr \left( \mathbf{A}^{-1} \sum_{k=1}^K \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right) \left( \mathbf{b}_k^{(3)} \right)^T \right] \right) \end{aligned} \quad (56)$$

Then, setting  $\mathbf{R} = \sum_{k=1}^K \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right) \left( \mathbf{b}_k^{(3)} \right)^T \right]$  simplifies to:

$$= -\frac{K}{2}\log|\mathbf{A}| - \frac{1}{2}tr(\mathbf{A}^{-1}\mathbf{R}) \quad (57)$$

Again, using Anderson and Olkin [218], the first derivative of this expression with respect to the elements of  $\mathbf{A}$  can be found. By setting equal to zero and rearranging the MLE for the covariance matrix of the study level random effects  $\mathbf{A}$  can be expressed.

$$\begin{aligned} \ell'(\boldsymbol{\Omega}) &= -\frac{K}{2}\mathbf{A}^{-1} - \frac{1}{2}\mathbf{A}^{-1}\mathbf{R}\mathbf{A}^{-1} \\ 0 &= -\frac{K}{2}\mathbf{A}^{-1} - \frac{1}{2}\mathbf{A}^{-1}\mathbf{R}\mathbf{A}^{-1} \\ \hat{\mathbf{A}} &= \frac{\mathbf{R}}{K} = \frac{\sum_{k=1}^K \mathbb{E} \left[ \left( \mathbf{b}_k^{(3)} \right) \left( \mathbf{b}_k^{(3)} \right)^T \right]}{K} \end{aligned} \quad (58)$$

#### 5.1.2.2.2.7 Additional procedures conducted by the function

Once both the E-step and the M-step in an iteration have been completed, the estimates of the model parameters given by the current and the previous iteration are compared. If the largest difference observed between any of the parameter estimates from consecutive iterations is less than the value supplied to the tolerance argument *tol* in the function call, then the iterative procedure ends, and the estimated model parameters are returned.

Once the model parameters have been estimated by the EM algorithm, the function calculates the log-likelihood of the model, by inputting the estimated model parameters into the log of the likelihood expression shown in equation (21). Again, the random effects are integrated out using a pseudo adaptive Gauss-Hermite procedure, as described in Section 5.1.2.2, with number of quadrature points set using the function argument *Lgpt*.

## 5.2 Simulation of multi-study joint data

The package contains a function to simulate multi-study joint longitudinal and time-to-event data, namely *simjointmeta()*. The methodology for simulation of data under a proportional hazards model with time varying covariates is described in Bender et al [219] and Austin [220]. In the following section, a general overview of the methods used to simulate data by the *simjointmeta* function is given. Details of data simulations conducted under specific conditions during later simulation studies are given in Chapter 7.

During discussion of the data simulation function, expressions are presented with a single individual level association parameter  $\alpha^{(2)}$  and a single study level association parameter  $\alpha^{(3)}$ . If separate association parameters were specified for each random effect (see Section 4.4.1) then these common association parameters would simply be replaced with multiple parameters e.g.  $\alpha_0^{(2)}, \alpha_1^{(2)}$ . If separate association parameters are specified for each study,  $\alpha^{(2)}$  and  $\alpha^{(3)}$  would be replaced with  $\alpha_k^{(2)}$  and  $\alpha_k^{(3)}$  for the case of common association parameter across random effects at the same level, or  $\alpha_{k0}^{(2)}, \alpha_{k1}^{(2)}$  and  $\alpha_{k0}^{(3)}, \alpha_{k1}^{(3)}, \alpha_{k2}^{(3)}$  for the case of separate association parameters for each random effect (similarly for parameters  $\theta_0, \theta_1$ , and, if supplied,  $\lambda$ , specified using arguments *theta0*, *theta1*, and *censLam* respectively). Expressions using common values are given, but note that the function allows study specific values for these arguments to be used.

The data simulation function in the package simulates data under a model that includes a fixed treatment assignment term in the time-to-event sub-model, and a fixed intercept, time and treatment term in the longitudinal sub-model. All data within a study is simulated at the same time. Initially, a full design matrix for the fixed effects is generated, so that when simulating study  $k$  (for  $k = 1 \dots K$  where  $K$  is the total number of studies) of size  $n_k$  individuals with  $m_k$  total possible longitudinal measurements, the longitudinal fixed effects design matrix for the study  $\mathbf{X}_{1k}$  is  $(n_k * m_k)$  by 3 (the number of fixed effects in the longitudinal sub-model), and the time-to-event design matrix  $\mathbf{X}_{2k}$  is  $n_k$  by 1. The treatment assignment variable is generated from a binomial  $B(n_k, 0.5)$  distribution, with a long version included in  $\mathbf{X}_{1k}$ . The design matrices for the individual level random effects  $\mathbf{Z}_k^{(2)}$  will be a subset of  $\mathbf{X}_{1k}$  containing only the covariates assigned individual level random effects, similarly for  $\mathbf{Z}_k^{(3)}$  if study level random effects are to be included.

To generate the individual level random effects for individuals in study  $k$ , labelled  $\mathbf{b}_k^{(2)}, n_k$  realizations from a zero mean multivariate normal distribution of dimension  $q$  (the number

of individual level random effects, equal to 2 if *rand\_ind* is set to “*intslope*”, or 1 if *rand\_ind* is set to “*int*”) are generated. If study level random effects, labelled  $\mathbf{b}^{(3)}$ , are specified,  $K$  realizations (one for each study  $k$  to be simulated) are generated from a multivariate distribution of dimension  $r$  (the number of study level random effects, equal to 2 if *rand\_stud* is set to “*inttreat*”, and 1 if *rand\_stud* is set to “*int*” or “*treat*”). The independent but identically distributed longitudinal measurement errors  $\boldsymbol{\varepsilon}_k$  are generated as a vector of realizations from a  $N(0, 1)$  distribution multiplied by the square root of the measurement error variance (specified by argument *vare*), with one realization for each simulated longitudinal measurement.

The longitudinal measurements  $\mathbf{Y}_k$  for study  $k$  for all individuals at all time points are then calculated using the following equation:

$$\mathbf{Y}_k = \mathbf{X}_{1k}\boldsymbol{\beta}_1 + \mathbf{Z}_k^{(2)}\mathbf{b}_k^{(2)} + \mathbf{Z}_k^{(3)}\mathbf{b}_k^{(3)} + \boldsymbol{\varepsilon}_k \quad (59)$$

In equation (59),  $\boldsymbol{\beta}_1$  are the population coefficients for the longitudinal covariates (specified using argument *beta1*). If study level random effects are not specified, then the  $\mathbf{Z}_k^{(3)}\mathbf{b}_k^{(3)}$  is not included in the model. At first, a longitudinal measurement for each individual within each study at each possible time point is generated. Measurements taken after the simulated event time are discarded before the data is returned by the function.

The distribution the time-to-event data is simulated under depends on the presence of time varying covariates in the random effects specification. No time-varying covariates are included as fixed effects in the model the time-to-event data is simulated under. However the joint data is simulated under a random effects only association structure, meaning that if data is simulated under a joint model containing a random time effect, time will be present in the time-to-event sub-model through the association structure. If time is not assigned a random effect, all terms in the time-to-event sub-model are time stationary. There is no option in the function to include study level random effects for time varying covariates (and so during this section the term  $\mathbf{Z}_k^{(3)}\mathbf{b}_k^{(3)}$  is not expanded), however time can be assigned an individual level random effect.

If time is present in the time-to-event sub-model (because *rand\_ind* = “*intslope*”) then the risk of hazard changes over time. As such, the function simulates data under a distribution that permits the baseline hazard to vary over time, namely the Gompertz distribution. If time is not present in the time-to-event sub-model (because *rand\_ind* = “*int*”) then the risk of hazard remains constant over time. As such, the function simulates

data under a distribution that holds the baseline hazard constant over time, namely the exponential distribution. This approach for simulating multi-study joint data is based on the theory for simulating time-to-event information with time varying covariates reported by Bender et al [219] and Austin [220], and the single study data simulation function contained in `joiner` [80].

A discussion of data simulated under a model with a random intercept at the individual level is now given, i.e. `rand_ind = "int"`. In this case, time-to-event data is simulated for study  $k$  under the following model:

$$\lambda_k(t) = \lambda_0(t) \exp\left(\beta_{21} \text{treat}_k + \alpha^{(2)} \mathbf{b}_{0k}^{(2)} + \alpha^{(3)} \mathbf{Z}_k^{(3)} \mathbf{b}_k^{(3)}\right) \quad (60)$$

In equation (60), the time-to-event sub-model contains a fixed treatment effect (where  $\text{treat}_k$  contains the treatment assignments for all individuals in study  $k$ ), a shared individual level random intercept with individual level association parameter  $\alpha^{(2)}$ , and shared time stationary study level random effects with study level association parameter  $\alpha^{(3)}$ . If study level random effects are not specified, then the  $\alpha^{(3)} \mathbf{Z}_k^{(3)} \mathbf{b}_k^{(3)}$  component will not appear in the model. In this case, as the time-to-event sub-model does not directly involve time varying covariates the event times are generated using the expression in equation (61) (see [219, 220]):

$$\mathbf{T}_{Ek} = \frac{-\log(\mathbf{U}_k)}{\exp\left(\theta_0 + \beta_{21} \text{treat}_k + \alpha^{(2)} \mathbf{b}_{0k}^{(2)} + \alpha^{(3)} \mathbf{Z}_k^{(3)} \mathbf{b}_k^{(3)}\right)} \quad (61)$$

In the above expression,  $\mathbf{T}_{Ek}$  is the vector of estimated event times for study  $k$ ,  $\mathbf{X}_{2k}$  is the stacked design matrix for the time-to-event sub-model for all  $n_k$  individuals in study  $k$ , and  $\beta_2$  is the coefficient for the time-to-event treatment effect as specified to argument `beta2`. Again, if study level random effects are not specified in the data simulation function call, the  $\alpha^{(3)} \mathbf{Z}_k^{(3)} \mathbf{b}_k^{(3)}$  component would not appear. The  $\theta_0$  parameter is the exponential of the  $\lambda$  parameter for an exponential distribution, specified in the function as argument `theta0`. The parameter  $\mathbf{U}_k$  is a vector of  $n_k$  realisations from  $U(0,1)$ .

In the methods presented in Bender et al [219], and Austin [220], the distribution of the baseline hazard that data is simulated under, and the distribution type of the event times, is the same. As such, the expression in equation (61) simulates event times under an exponential distribution, and the baseline hazard is exponentially distributed. This is because the time-to-event sub-model does not involve time-varying covariates, and as

noted in Bender et al [219], the baseline hazard for an exponential distribution is a constant value: it does not vary with time.

Alternatively, if data is simulated under a model with both a random intercept and a random time term at the individual level, i.e. *rand\_ind = "intslope"*, then the time-to-event sub-model data is simulated under has specification:

$$\lambda_k(t) = \lambda_0(t) \exp(\beta_{21} \text{treat}_k + \alpha^{(2)}(b_{0k}^{(2)} + b_{1k}^{(2)} \text{time}_k) + \alpha^{(3)} Z_k^{(3)} b_k^{(3)}) \quad (62)$$

Equation (62), has the same parameters as equation (60), with the addition of an individual level random slope  $b_{1k}^{(2)} \text{time}_k$ . As before,  $\alpha^{(3)} Z_k^{(3)} b_k^{(3)}$  will only be included in the model if study level random effects are specified in the function call. Unlike before, data are now being simulated under a time-to-event model with a time varying covariate present in the time-to-event sub-model through the association structure. As such, simulating event times under an exponential distribution is no longer appropriate, as the baseline hazard will not remain constant over time. Instead, event times are simulated under a Gompertz distribution (see [219, 220]), which is often used to represent human lifetimes. The Gompertz distribution has a baseline hazard that can vary over time, which is necessary given that equation (62) contains a time varying random effect  $b_{1k}^{(2)}$ . As such, event times are generated under the equation:

$$T_{Ek} = \frac{1}{\alpha^{(2)} b_{1k}^{(2)} + \theta_1} \log \left[ 1 + \frac{(\alpha^{(2)} b_{1k}^{(2)} + \theta_1)(-\log(U_k))}{\exp(\theta_0 + \beta_{21} \text{treat}_k + \alpha^{(2)} b_{0k}^{(2)} + \alpha^{(3)} Z_k^{(3)} b_k^{(3)})} \right] \quad (63)$$

In equation (63),  $U_k$  again represents  $n_k$  realisations from  $U(0, 1)$ ,  $\alpha^{(2)}$  and  $\alpha^{(3)}$  are respectively the individual level and study level association parameters,  $b_{0k}^{(2)}$  are the individual level random intercepts and  $b_{1k}^{(2)}$  the individual level random time terms for study  $k$ , and  $Z_k^{(3)} b_k^{(3)}$  are the study level random effects and their design matrix. The parameters  $\theta_0$  (the exponential of which is the scale parameter of a Gompertz distribution, i.e.  $\kappa = \exp \theta_0$ ) and  $\theta_1$  (the shape parameter of a Gompertz distribution) are used along with the coefficients in the model to control the distribution of the event times. They are specified in the function call through arguments *theta0* and *theta1*. Bender et al [219] suggested using the extreme value theory to calculate values to set *theta0* and *theta1* to, given target mean and variance for the event times.

A Gompertz distribution has increasing hazard for a positive shape parameter, constant hazard for a shape parameter equal to 0 (equivalent to an exponential distribution), and a

decreasing hazard for negative shape parameters. Under the above model, the probability density function of the event times takes form:

$$f_0(t) = \kappa \exp(\theta_1 t) \exp\left(\frac{\kappa}{\theta_1} (1 - \exp(\theta_1 t))\right), \text{ where } \kappa = \exp \theta_0 \quad (64)$$

However, if the shape parameter is negative, if time is allowed to tend towards infinity, there is a non-zero probability of living forever. As such, in the function, event times when the Gompertz distribution is employed are simulated under a two-step process. First, for each individual  $i$  within study  $k$ , the following two conditions are checked (using the uniform realizations held in  $\mathbf{U}_k$ ).

$$\begin{aligned} \text{Condition 1: } & (\theta_1 + \alpha^{(2)} b_{1ki}^{(2)}) < 0 \\ \text{Condition 2: } & U_{ki} < \exp\left(\frac{\exp(\theta_0 + \alpha^{(2)} b_{0ki}^{(2)})}{\theta_1 + \alpha^{(2)} b_{1ki}^{(2)}}\right) \end{aligned}$$

If the conditions are both true, the individual is automatically assigned an event time of infinity, otherwise their event time is generated under equation (63). As such, unless it is acceptable for individuals to be assigned infinite event times when *rand\_ind* = “*intslope*”, it is important to either allow censoring, or to specify that event times be truncated at a certain time (using arguments *trunctation* and *trunctime*).

Regardless of the individual level random effects specification that the data is generated under, if the data is specified as censored (i.e. *censoring* = *TRUE*), the censoring time is generated in the same way. The censoring times are always exponentially distributed, under  $\exp(\lambda)$ , where  $\lambda$  is specified in the function call using argument *censLam*. As such, censoring times  $T_{ck}$  for study  $k$  are generated using equation:

$$T_{ck} = \frac{-\log(\mathbf{U}_k)}{\lambda} \quad (65)$$

In equation (65), as earlier,  $\mathbf{U}_k$  is a vector of length  $n_k$  of realizations from  $U(0,1)$ , and  $\lambda$  is the value of *censLam* as specified in the function call. If censoring is permitted, then the survival time returned by the distribution is the minimum of the generated censoring and event times, with, for each individual, the censoring variable taking value 1 if  $T_{Eki} < T_{Cki}$ , 0 otherwise.

Once both the longitudinal and time-to-event data has been simulated in this manner, the simulated joint dataset is returned in the format described in Section 4.4.1.

### 5.3 Discussion

During this chapter the methodology which underpins the functions made available in the multi-study joint modelling package *joinerMeta* has been presented and discussed. The theory behind multi-study joint longitudinal and time-to-event data simulation has been presented, as well as the background as to why certain association structures cause effects on event risks to be decomposable into direct and indirect effects. The methods used to fit one-stage multi-study joint models have been described, and areas of future work to either increase or improve the current model fitting provided by the package have been summarized.



## Chapter 6 : Real Data Applications of Joint Meta-Analytic Methods

---

In this chapter, the one-stage and two-stage meta-analytic methods for joint data described in Chapter 3 are applied to a real dataset. Issues and considerations of the methods when used in an applied research example will be discussed, and the results motivate the simulation studies conducted in Chapter 7. The chapter will conclude with a discussion of the results of the real world example.

### 6.1 Description of the INDANA dataset

The INDANA dataset produced by the INDANA collaboration [91] is an IPD dataset comprising of data for hypertensive (high blood pressure) patients from multiple studies. It was assembled to determine how the efficacy of pharmacological treatment for high blood pressure depended on patient characteristics [91]. The investigation included any randomised controlled trials that assessed the efficacy of any drug interventions for hypertension (denoted by treatment group 1), versus no treatment, placebo or usual care (denoted by treatment group 0). A total of 14 trials were included in the collaboration. During this research, a subset of the INDANA dataset was examined, which included any studies containing both longitudinal and time-to-event IPD. The subset analysed (henceforward referred to as the INDANA data) consisted of data from 6 trials [221-226]. The individual studies in the subset are referred to as EWPHE [221], COOP [222], STOP [223], SHEP [224], MRC1 [225] and MRC2 [226] respectively.

The dataset contained two continuous longitudinal outcomes: systolic and diastolic blood pressure (denoted SBP and DBP respectively). Three time-to-event outcomes are available; time to death, time to myocardial infarction (MI, death of a portion of heart muscle due to blood supply disruption resulting in heart attack [227]), and time to stroke (a sudden attack of weakness affecting one side of the body caused by interrupted blood supply to the brain [227]). The latter two time-to-event outcomes included both terminal and non-terminal events. These outcomes are all linked to cardiovascular disease (CVD) [228]. As this thesis is concerned with univariate joint models, analysis is restricted to three pairs of outcomes, namely SBP and time to death, SBP and time to MI, and SBP and time to stroke. SBP was analysed rather than DBP as SBP is noted to be a more significant issue as age increases,

and elevated DBP is more common in populations below 50 [229]. As such, given the demographic of the INDANA dataset (Section 6.1.1), an analysis of SBP is more appropriate.

For the EWPHE trial an intention to treat analysis is only possible for fatal endpoints, and so the study contributed data only to the SBP and time to death analysis. The final joint dataset used in this thesis contained a maximum of 6 studies totalling at most 29837 individuals. Exact numbers of individuals contributing to each analysis are available in each analysis' results table.

During these analyses, the aim was to demonstrate methods developed during this thesis rather than to investigate potential treatment modifiers. As such, although the INDANA datasets contained additional patient covariates, models in this investigation only involved the treatment assignment and longitudinal measurement time covariates. In addition, as analyses are based on a subset of available studies (containing those that could provide both longitudinal and time-to-event IPD), any clinical interpretation should be made with caution.

#### 6.1.1 Comparison of demographics between studies

Heterogeneity between studies can occur due to differences in the patients' demographics. For example, the observed treatment effect may differ between studies with elderly compared to younger populations. The INDANA dataset contains a range of baseline variables measured in some or all the included studies (data in Table 6 and Table 7, levels of missing data in Table 8). Many of these recorded covariates are known to influence CVD (Cardiovascular Disease) or blood pressure, for example many of them are included in the QRISK model (which produces a CVD risk score), or its update QRISK2 (which includes the variables in QRISK, as well as several additional variables such as patient reported ethnicity) [230-235]. The demographic characteristics in the INDANA dataset show potential differences between the studies, which, for covariates linked to CVD, may cause between study heterogeneity.

The size of the studies included in the INDANA data varies considerably, with the smallest study contributing 840 individuals (EWPHE), and the largest containing 17354 individuals (MRC1, over half of the IPD). The proportions of individuals assigned to each treatment group is similar across studies (close to equal proportions). Table 6 shows the discrete baseline measurements available in the INDANA dataset, from which several areas can be identified where the demographics differ between studies.

Firstly, for most studies, between 60% and 70% of the study population are female. However, the largest study in the meta-analysis, MRC1, had closer to equal gender proportions. Within each study, the proportion of females in each treatment group were comparable, as expected due to the randomisation process. As discussed in Reckelhoff [236], age-matched men are generally at higher risk of CVD than premenopausal women, and women had lower 24 hour mean blood pressure than men until the age of 70-79 years. Additionally, gender is included in the aforementioned QRISK2 model as potentially related to CVD [231, 232, 235]. The difference between proportion of females between MRC1 and the other studies could result in heterogeneity.

The proportion of smokers varies at baseline between studies, with only 7.7% of those in the STOP trial classed as a smoker at baseline, compared to 28.9% in the MRC1 study. Within studies, there were similar proportions of smokers between treatment groups, apart from COOP which had 16.8% smokers at baseline in the no treatment, placebo or usual care group, and 22.7% in the any drug intervention group. Smoking is a risk factor for CVD [235], with NICE guidelines for hypertension in adults recommending support to help patients stop smoking [229]. Consequently, the variation between studies in proportion of smokers may lead to between study heterogeneity, and the inconsistency between treatment groups in COOP may reduce apparent treatment benefit.

There was a disparity in proportion of individuals with history of diabetes between studies. Some trials (COOP, MRC1, MRC2) contained no, or close to zero, individuals with history of diabetes, whereas the remaining studies (EWPHE, SHEP, STOP) involved populations where just under 10% had a history of diabetes. Proportions of individuals with history of diabetes were comparable between treatment groups within studies. Type 2 diabetes is included in the QRISK2 model [231, 232, 235] used to predict CVD. Whilst information on type of diabetes was not provided in the dataset, variability in proportions of those with history of diabetes between study populations could lead to between study heterogeneity.

History of treatment for high blood pressure is a missing variable for the STOP trial. However, where recorded, there is disparity between studies, with COOP, MRC1 and MRC2 containing no individuals with history of treatment for high blood pressure, whereas EWPHE contains 57.7%, and SHEP contains 33.3%. The proportions between treatment groups are similar for SHEP, whereas slightly fewer in the any drug intervention group in EWPHE have a history of treatment for high blood pressure (55% compared to 60.4%). History of high blood pressure was also included in the dataset, but this was only recorded

in the MRC1 trial, who recorded that none of patients had a history of high blood pressure. The QRISK2 model includes previous treatment for hypertension as a variable, indicating a possible link to CVD. As such, again the treatment response between studies could differ due to the differing populations.

The dataset also includes variables for history of MI (myocardial infarction) and history of stroke. The proportions were low in most cases, with only EWPHE reporting higher proportions than 5% for history of stroke in both treatment groups. In general, proportions with history of each event were similar across studies and between treatment groups within studies.

Table 7 contains summaries of the continuous variables included in the dataset. Again many are linked to CVD. As mentioned earlier, increasing age is linked with increasing risk of hypertension [234]. Most studies in the meta-analysis have a mean age around 70 years old, apart from MRC1, which has a mean age of 52.1 years. Within each study the mean age was similar between treatment groups. As hypertension depends on age, as evidenced by the covariate's inclusion in the QRISK and QRISK2 models [234, 235], the large difference in mean age between MRC1 and the other studies could manifest as between study heterogeneity.

Height and weight are included in the INDANA dataset, and Body Mass Index (BMI, derived from these measurements) is known to affect CVD [230, 234, 235]. As such, BMI was calculated and compared. Mean BMI within each study was comparable between treatment groups. Across studies, BMI was again comparable, with largest mean BMI of 27.6 in SHEP, and lowest of 26.4 in EWPHE. Consequently, BMI, whilst having the potential to affect the outcomes of this real example, may not contribute greatly to between study heterogeneity in this case.

Baseline heart rate is also available in the dataset. The effect of heart rate on hypertension has been widely debated [237, 238], and it is not commonly used as a predictor for CVD. This value was unmeasured in the EWPHE and MRC1 trials, and showed some variability between trials where it was recorded (e.g. mean baseline heart rate in SHEP was 70.8, whilst in MRC2 it was 82.2). Within studies, baseline heart rate was comparable within treatment groups. However, as a link between CVD or hypertension and heart rate has not been definitely established, this variability may not cause or explain between study heterogeneity.

Baseline was included as a time-point in longitudinal analyses, and so baseline SBP constituted part of the outcome for the longitudinal models. Baseline DBP was also included in the dataset. Variability between studies existed for SBP, with MRC2 giving a mean baseline SBP of 161.6, whilst the COOP trial reported a mean baseline SBP OF 197.1. A diagnosis of hypertension can stem from elevated systolic or diastolic blood pressure, although elevated DBP is more common for those under 50, whilst SBP is a greater issue for older populations [229]. Baseline DBP varied between studies, with lowest mean value 76.8 in SHEP, whilst highest in STOP with 101.7. There was little variability within studies between treatment groups. Such variability at baseline in SBP or DBP is likely to translate into variability between studies in treatment estimates.

Serum cholesterol is a recognised predictor for CVD [234], however mean baseline cholesterol was similar between studies and between treatment groups within studies, and so was unlikely to cause severe heterogeneity. The dataset also contained baseline serum creatinine, which can be used to measure renal function (important as hypertension is a risk factor for chronic kidney disease [239]). This variable was not recorded in the MRC1 trial, and was similar for all other trials apart from SHEP, which reported marginally higher mean values (within studies mean values were comparable between treatment groups). The differences in this variable between studies could lead to heterogeneity, however the variable is a measure of the effect of hypertension or its treatment on renal performance, and so may not directly affect the outcomes of interest.

For both the discrete and continuous data, some variables contained missing data (Table 8). In most cases where missing data was present, data was available for the covariate for over 95% of the population. However, some variables displayed much higher levels of missingness (e.g. baseline heart rate was missing for 62% of the population), and several variables were completely unrecorded in some studies (identified by NR=Not Recorded). This does not severely affect this investigation as the covariates are not used, but it could present an issue for this disease area in the future if covariates with a high level of missingness are found to significantly affect the outcomes of interest.

Study	Treatment group (%)	Number Female (%)	Baseline smoking status - smoker (%)	History of Diabetes (%)	History of treatment for High Blood Pressure (%)	History of High Blood Pressure (%)	History myocardial infarction (%)	History of stroke (%)
COOP	0	465 (52.6)	314 (67.5)	78 (16.8)	0 (0)	0 (0)	10 (2.2)	5 (1.1)
	1	419 (47.4)	297 (70.9)	95 (22.7)	0 (0)	0 (0)	9 (2.1)	8 (1.9)
	Total	884	611 (69.1)	173 (19.6)	0 (0)	0 (0)	19 (2.1)	13 (1.5)
EWPHE	0	424 (50.5)	299 (70.5)	67 (15.8)	35 (8.3)	256 (60.4)	0 (0)	23 (5.4)
	1	416 (49.5)	287 (69.0)	76 (18.3)	37 (8.9)	229 (55.0)	0 (0)	30 (7.2)
	Total	840	586 (69.8)	143 (17.0)	72 (8.6)	485 (57.7)	0 (0)	53 (6.3)
MRC1	0	8654 (49.9)	4129 (47.7)	2531 (29.2)	1 (0.01)	0 (0)	41 (0.5)	0 (0)
	1	8700 (50.1)	4177 (48.0)	2488 (28.6)	4 (0.05)	0 (0)	46 (0.5)	0 (0)
	Total	17354	8306 (47.9)	5019 (28.9)	5 (0.03)	0 (0)	87 (0.5)	0 (0)
MRC2	0	2213 (50.3)	1287 (58.2)	485 (21.9)	0 (0)	0 (0)	0 (0)	0 (0)
	1	2183 (49.7)	1273 (58.3)	476 (21.8)	0 (0)	0 (0)	0 (0)	0 (0)
	Total	4396	2560 (58.2)	961 (21.9)	0 (0)	0 (0)	0 (0)	0 (0)
SHEP	0	2371 (50.1)	1359 (57.3)	305 (12.9)	241 (10.2)	794 (33.5)	95 (4.0)	40 (1.7)
	1	2365 (49.9)	1331 (56.3)	297 (12.6)	237 (10.0)	781 (33.0)	87 (3.7)	59 (2.5)
	Total	4736	2690 (56.8)	602 (12.7)	478 (10.1)	1575 (33.3)	182 (3.8)	99 (2.1)
STOP	0	815 (50.1)	509 (62.5)	61 (7.5)	60 (7.4)	-	16 (2.0)	35 (4.3)
	1	812 (49.9)	510 (62.8)	64 (7.9)	73 (9.0)	-	17 (2.1)	31 (3.8)
	Total	1627	1019 (62.6)	125 (7.7)	133 (8.2)	-	33 (2.0)	66 (4.1)
Overall	0	14942 (50.1)	7897 (50.1)	3527 (23.6)	337 (2.3)	1050 (7.0)	162 (1.1)	103 (0.7)
	1	14895 (40.9)	7875 (49.9)	3496 (23.5)	351 (2.4)	1010 (6.8)	159 (1.1)	128 (0.9)
	Total	29837	15772 (52.9)	7023 (23.5)	688 (2.3)	2060 (6.9)	321 (1.1)	231 (0.8)

Table 6: Discrete Baseline measurements overall, by study and by treatment group within study (with treatment group 0 representing placebo, no treatment or usual care, treatment group 1 representing assignment to any drug intervention for hypertension).

Study	Treatment group (%)	Mean Age in years (sd)	Mean Baseline		Mean Baseline Heart Rate in beats per minute (sd)		Mean Baseline SBP in mmHg (sd)		Mean Baseline DBP in mmHg (sd)		Mean Baseline cholesterol in mmol/L (sd)		Mean Baseline creatinine levels in μmol/L (sd)	
			BMI in kg/m² (sd)	BMI in kg/m² (sd)	Heart Rate in beats per minute (sd)	Heart Rate in beats per minute (sd)	SBP in mmHg (sd)	SBP in mmHg (sd)	DBP in mmHg (sd)	DBP in mmHg (sd)	cholesterol in mmol/L (sd)	cholesterol in mmol/L (sd)	creatinine levels in μmol/L (sd)	creatinine levels in μmol/L (sd)
COOP	0	465 (52.6)	69.1 (5.1)	26.5 (5.1)	81.5 (14.1)	196.5 (20.3)	97.4 (13.4)	6.9 (1.5)	87.2 (18.8)					
	1	419 (47.4)	69.0 (5.1)	26.4 (4.7)	82.2 (13.5)	197.7 (21.4)	99.9 (14.4)	6.9 (1.4)	88.2 (18.0)					
	Total	884	69.0 (5.1)	26.5 (4.9)	81.9 (13.8)	197.1 (20.8)	98.6 (13.9)	6.9 (1.5)	87.7 (18.4)					
EWPHE	0	424 (50.5)	71.9 (8.1)	26.5 (4.5)	-	182.4 (16.0)	100.5 (7.1)	6.4 (1.4)	89.2 (22.3)					
	1	416 (49.5)	71.6 (8.0)	26.3 (4.6)	-	182.9 (16.9)	100.6 (7.1)	6.3 (1.3)	88.9 (22.3)					
	Total	840	71.8 (8.0)	26.4 (4.5)	-	182.6 (16.5)	100.5 (7.1)	6.4 (1.3)	89.1 (22.3)					
MRC1	0	8654 (49.9)	52.1 (7.5)	27.0 (4.1)	-	161.6 (17.0)	98.3 (5.7)	6.4 (1.1)	-					
	1	8700 (50.1)	52.1 (7.5)	27.1 (4.2)	-	161.6 (17.1)	98.5 (5.7)	6.5 (1.1)	-					
	Total	17354	52.1 (7.5)	27.1 (4.1)	-	161.6 (17.1)	98.4 (5.7)	6.5 (1.1)	-					
MRC2	0	2213 (50.3)	70.3 (2.8)	26.6 (4.0)	82.5 (12.8)	184.4 (13.0)	90.8 (11.7)	6.4 (1.2)	90.4 (18.8)					
	1	2183 (49.7)	70.3 (2.8)	26.5 (4.0)	81.8 (12.2)	184.8 (12.9)	90.7 (11.4)	6.5 (1.3)	90.9 (19.5)					
	Total	4396	70.3 (2.8)	26.5 (4.0)	82.2 (12.5)	184.6 (13.0)	90.8 (11.5)	6.5 (1.2)	90.6 (19.2)					
SHEP	0	2371 (50.1)	71.5 (6.7)	27.6 (5.1)	71.3 (10.5)	170.1 (9.2)	76.7 (8.6)	6.1 (1.1)	104.1 (22.9)					
	1	2365 (49.9)	71.6 (6.7)	27.6 (4.9)	70.3 (10.5)	170.5 (9.5)	76.9 (8.9)	6.1 (1.2)	104.7 (23.5)					
	Total	4736	71.6 (6.7)	27.6 (5.0)	70.8 (10.5)	170.3 (9.4)	76.8 (8.7)	6.1 (1.1)	104.4 (23.2)					
STOP	0	815 (50.1)	75.7 (3.7)	26.5 (3.8)	75.9 (10.6)	195.3 (13.7)	101.7 (7.2)	6.5 (1.3)	89.2 (19.1)					
	1	812 (49.9)	75.6 (3.7)	26.7 (3.9)	76.4 (10.6)	194.5 (14.1)	101.6 (7.0)	6.4 (1.2)	88.5 (19.2)					
	Total	1627	75.7 (3.7)	26.6 (3.8)	76.2 (10.6)	194.9 (13.9)	101.7 (7.1)	6.5 (1.2)	88.8 (19.1)					
Overall	0	14942 (50.1)	60.2 (11.7)	27.0 (4.3)	76.9 (12.8)	169.8 (19.2)	94.0 (11.2)	6.4 (1.2)	95.2 (21.9)					
	1	14895 (49.9)	60.2 (11.7)	27.0 (4.3)	76.3 (12.6)	169.8 (19.3)	94.2 (11.2)	6.4 (1.2)	95.6 (22.5)					
	Total	29837	60.2 (11.7)	27.0 (4.3)	76.6 (12.7)	169.8 (19.3)	94.1 (11.2)	6.4 (1.2)	95.4 (22.2)					

Table 7: Continuous baseline measurements overall, by study and by treatment group within study (with treatment group 0 representing placebo, no treatment or usual care, treatment group 1 representing assignment to any drug intervention for hypertension)

Number missing values (%)							Overall
Variable	COOP	EWPHF	MRC1	MRC2	SHEP	STOP	
Smoking Status	7 (0.8)	0 (0.0)	77 (0.4)	2 (0.0)	2 (0.0)	19 (1.2)	107 (0.4)
History of Diabetes	0 (0.0)	0 (0.0)	2 (0.0)	0 (0.0)	86 (1.8)	81 (5.0)	169 (0.6)
History of treatment for high blood pressure	0 (0.0)	9 (1.1)	0 (0.0)	0 (0.0)	0 (0.0)	NR: 1627	1636 (5.5)
History of high blood pressure	NR: 884	NR: 840	0 (0.0)	NR: 4396	NR: 4736	NR: 1627	12483 (41.8)
History of myocardial infarction	17 (1.9)	0 (0.0)	2 (0.0)	0 (0.0)	59 (1.2)	13 (0.8)	91 (0.3)
History of stroke	17 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	32 (0.7)	13 (0.8)	62 (0.2)
Baseline BMI	7 (0.8)	40 (4.8)	18 (0.1)	5 (0.1)	61 (1.3)	43 (2.6)	174 (0.6)
Baseline heart rate	8 (0.9)	NR: 840	NR: 17354	61 (1.4)	4 (0.1)	23 (1.4)	18290 (61.3)
Baseline systolic blood pressure	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.0)	0 (0.0)	15 (0.9)	17 (0.1)
Baseline diastolic blood pressure	0 (0.0)	0 (0.0)	0 (0.0)	9 (0.2)	15 (0.3)	15 (0.9)	39 (0.1)
Baseline cholesterol	120 (13.6)	18 (2.1)	388 (2.2)	21 (0.5)	370 (7.8)	650 (40.0)	1567 (5.3)
Baseline creatinine levels	303 (34.3)	11 (1.3)	NR: 17354	263 (6.0)	300 (6.3)	22 (1.4)	18253 (61.2)

Table 8: Number of missing values (% missing) from baseline variables by study and overall. Unlisted variables contained no missing data. NR = Not Recorded in the study.

### 6.1.2 Longitudinal Data

Possible measurement times for the longitudinal outcomes were baseline, 6 months, 1 year and annually thereafter to a maximum of 7 years (giving 9 potential measurement times), however measurement schedules differed between studies. The SHEP study recorded at least some individuals at 6 measurement times, and STOP and MRC1 at 7 measurement times. The remaining studies reported measurements at all 9 possible measurement times. Only an individual's longitudinal measurements made before their recorded survival time



contributed towards each analysis. The number of longitudinal measurements at each time point are available in Table 9.

<i>Study</i>	<i>Time (years)</i>								
	<i>0</i>	<i>0.5</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
<b><i>SBP and time to death</i></b>									
<b><i>COOP</i></b>	884	760	785	722	514	329	267	200	160
<b><i>EWPHE</i></b>	840	749	653	509	383	297	213	118	63
<b><i>MRC1</i></b>	17354	16525	16343	15308	14611	12584	8353	0	0
<b><i>MRC2</i></b>	4394	4182	4100	3765	3490	3223	2596	655	52
<b><i>SHEP</i></b>	4736	0	4243	4091	3938	2644	1164	0	0
<b><i>STOP</i></b>	1612	1520	1440	798	311	67	29	0	0
<b><i>SBP and time to MI</i></b>									
<b><i>COOP</i></b>	884	759	782	713	506	319	260	195	155
<b><i>MRC1</i></b>	17354	16512	16309	15253	14520	12478	8273	0	0
<b><i>MRC2</i></b>	4394	4176	4086	3739	3465	3183	2564	644	52
<b><i>SHEP</i></b>	4728	0	4220	4051	3877	2592	1141	0	0
<b><i>STOP</i></b>	1612	1518	1433	784	299	65	28	0	0
<b><i>SBP and time to stroke</i></b>									
<b><i>COOP</i></b>	884	754	777	709	496	316	251	187	153
<b><i>MRC1</i></b>	17354	16521	16325	15282	14572	12542	8318	0	0
<b><i>MRC2</i></b>	4394	4172	4080	3737	3451	3177	2547	642	51
<b><i>SHEP</i></b>	4736	0	4206	4005	3834	2555	1105	0	0
<b><i>STOP</i></b>	1612	1515	1418	767	294	66	29	0	0

Table 9: Number of longitudinal measurements available at each time point by study for analysis of SBP and time to death, SBP and time to MI, and SBP and time to stroke.

Plotting the longitudinal trajectories for SBP each study indicated a change in mean study population trajectory slope at approximately 6 months (Figure 37-Figure 39, Appendix 2). To account for this characteristic of the data, a variety of functions of the longitudinal time variable  $t$  were tested for inclusion in joint models fitted to each study and to the multi-study meta-dataset, (including  $t^2$  and  $\exp(-g * t)$ , where  $g$  is some constant), and the model fits compared using the deviance and the AIC scores. It was determined that the change in the longitudinal trajectory was well represented across the studies by including an  $\exp(-3 * t)$  term when modelling SBP. This approach is employed both in the two-stage and the one-stage investigations shown in this chapter.

The option to fit a changepoint model to the data was considered, and preliminary code to fit such a model developed. However, there were some issues with model convergence, and as the model containing an  $\exp(-3 * t)$  term fitted the data adequately, completion of the changepoint model code was delayed until a later date.

Figure 4-Figure 6 show the longitudinal trajectories panelled by event type (0 = censored, 1 = experienced the event in question), with longitudinal time adjusted by survival time ( $t_{kij} - T_{ski}$ ). This adjustment allows changes in the longitudinal trajectory just before the individual is censored or experiences the event to be identified. In many of the studies, for each of the time-to-event outcomes, the SBP values taken by those experiencing an event, compared to those censored, tend to be higher. Additionally, the loess smoother for those censored shows an initial decrease followed by a plateau, whilst those experiencing the event showed a loess smoother with a more constant linear decline. These differences in loess smoothers indicate a potential benefit of jointly modelling SBP and each of the three time-to-event outcomes of interest.

Panelling the graphs further by treatment assignment (Figure 7-Figure 9) additionally allows differences between those assigned to any drug intervention for hypertension (TREAT = 1) versus those assigned to no treatment, placebo or usual care (TREAT = 0) in SBP trajectories to be assessed. The SBP values for those assigned to any drug intervention appear marginally lower than those assigned to no treatment, placebo or usual care. The shapes of the loess smoothers remained relatively similar between treatment groups for those censored and also for those experiencing an event. Consequently, whilst assignment to any drug intervention for hypertension is likely to reduce SBP, the association between SBP and each of the time-to-event outcomes may not be significant.

## SBP and Time to Death

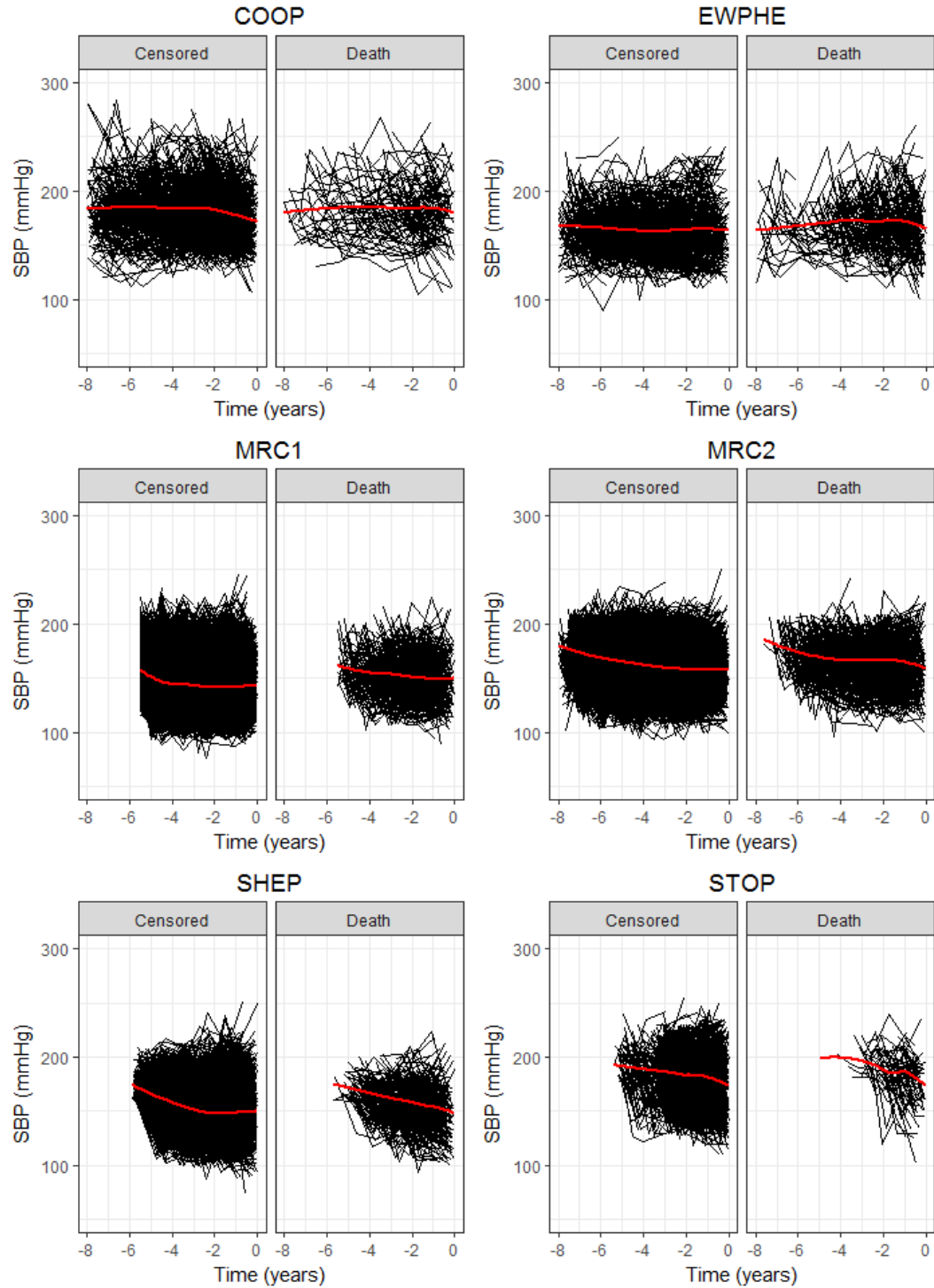


Figure 4: For SBP and time to death, individual longitudinal trajectories (black) panelled by event type for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).

## SBP and Time to MI

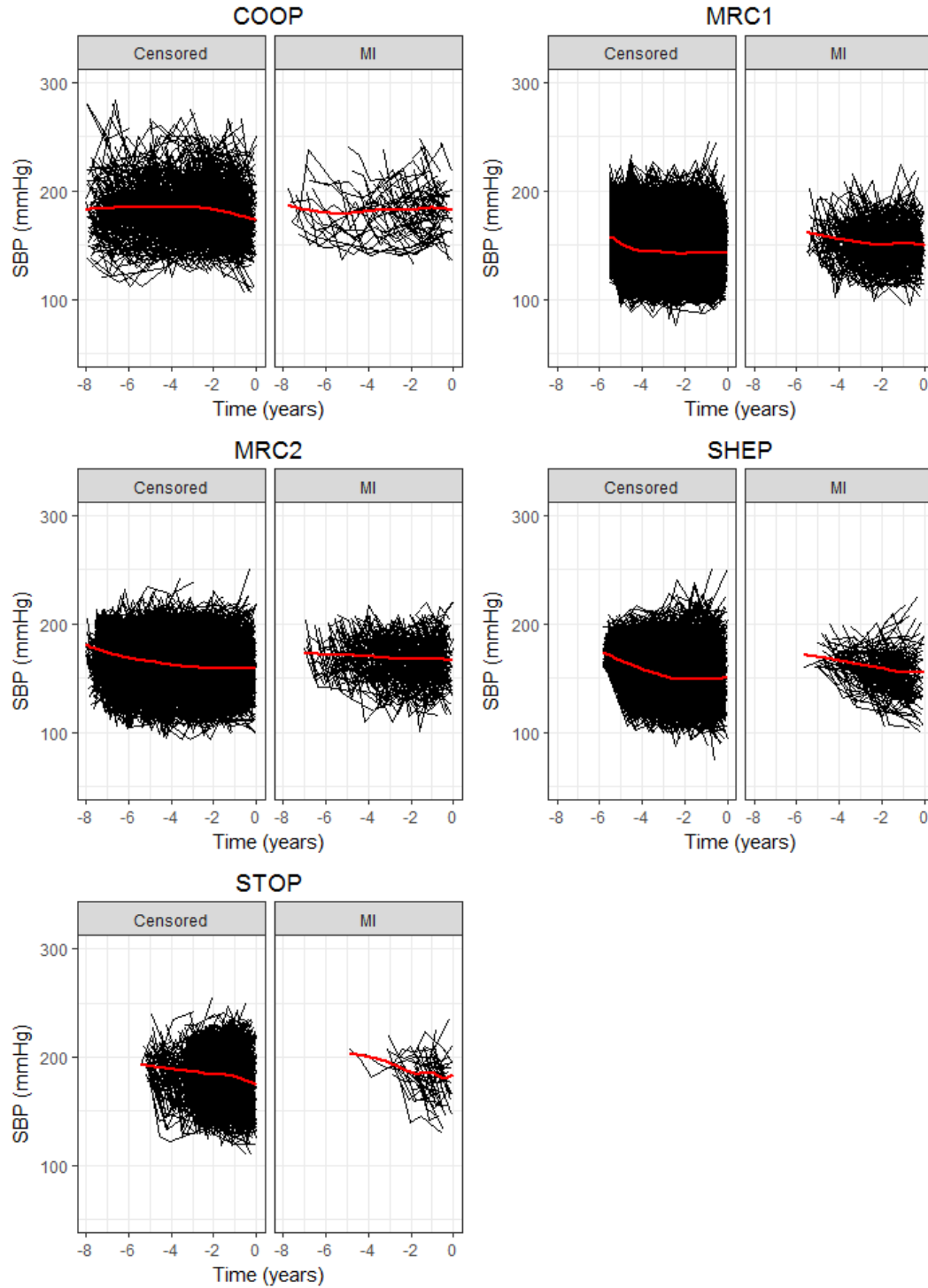


Figure 5: For SBP and time to MI data, individual longitudinal trajectories (black) panelled by event type for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).

## SBP and Time to Stroke

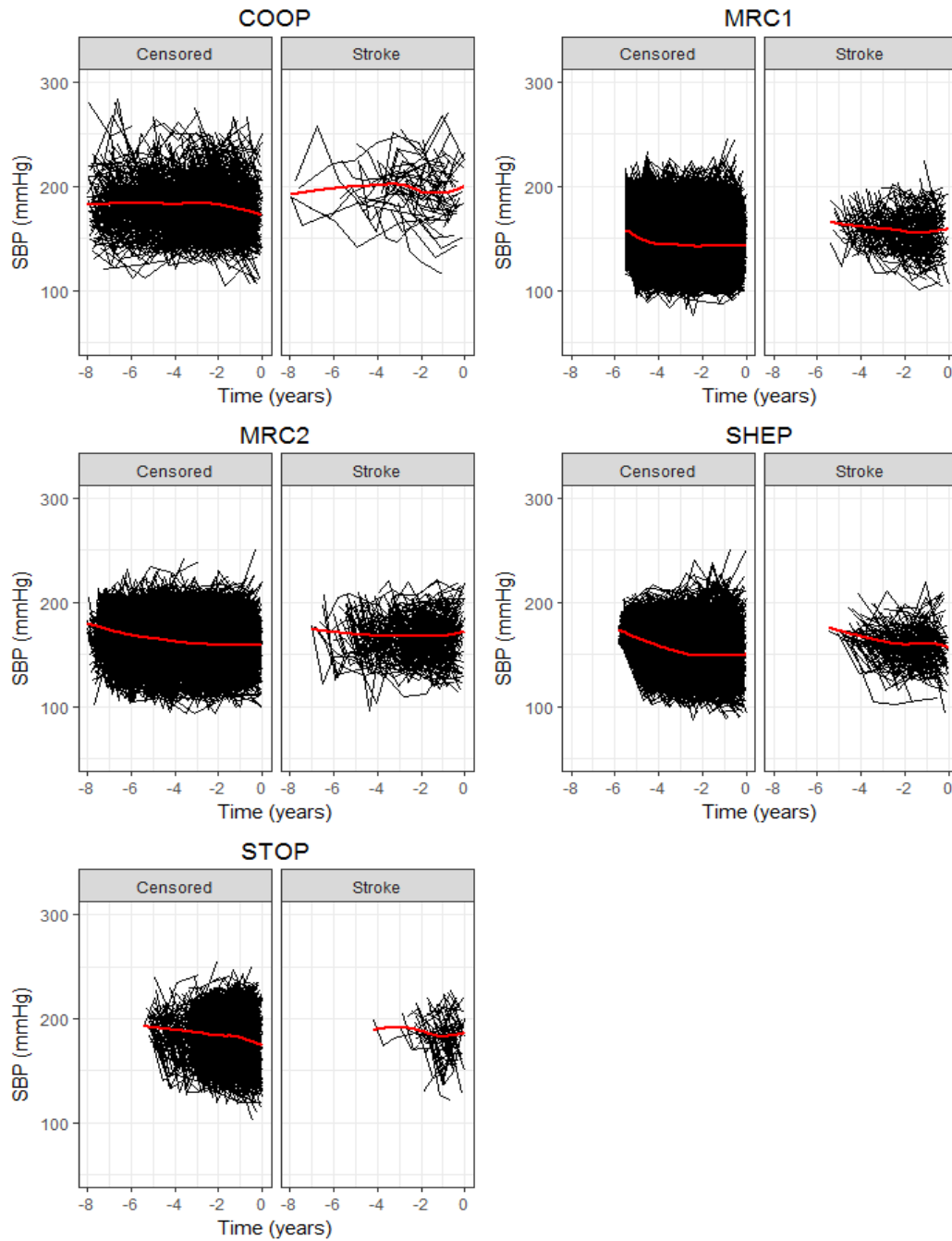


Figure 6: For SBP and time to stroke data, individual longitudinal trajectories (black) panelled by event type for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).

## SBP and Time to Death

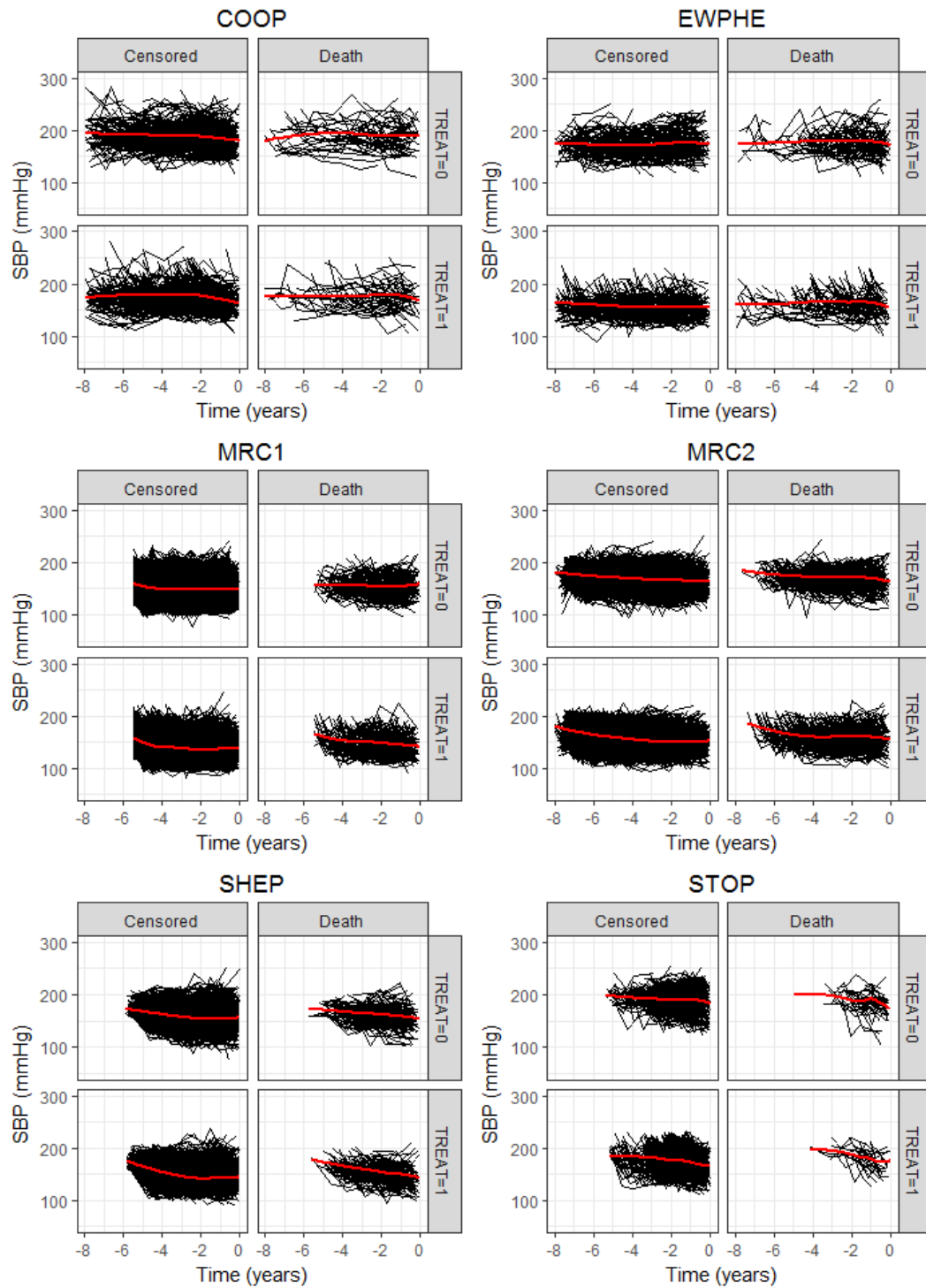


Figure 7: For SBP and time to death, individual longitudinal trajectories (black) panelled by event type and treatment group (1=any drug intervention for hypertension, 0= no treatment, placebo or usual care) for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).

## SBP and Time to MI

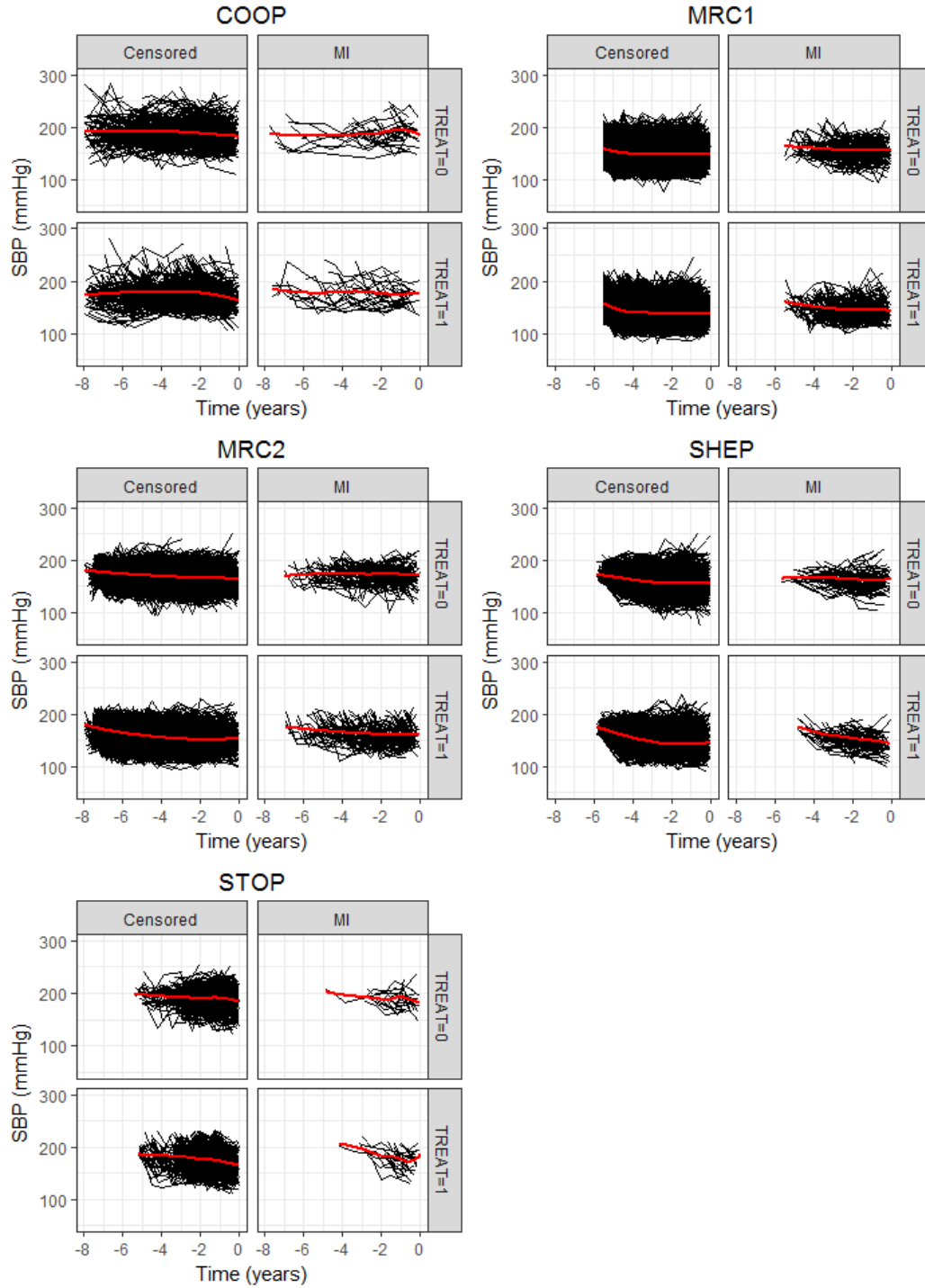


Figure 8: For SBP and time to MI data, individual longitudinal trajectories (black) panelled by event type and treatment group (1=any drug intervention for hypertension, 0= no treatment, placebo or usual care) for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).

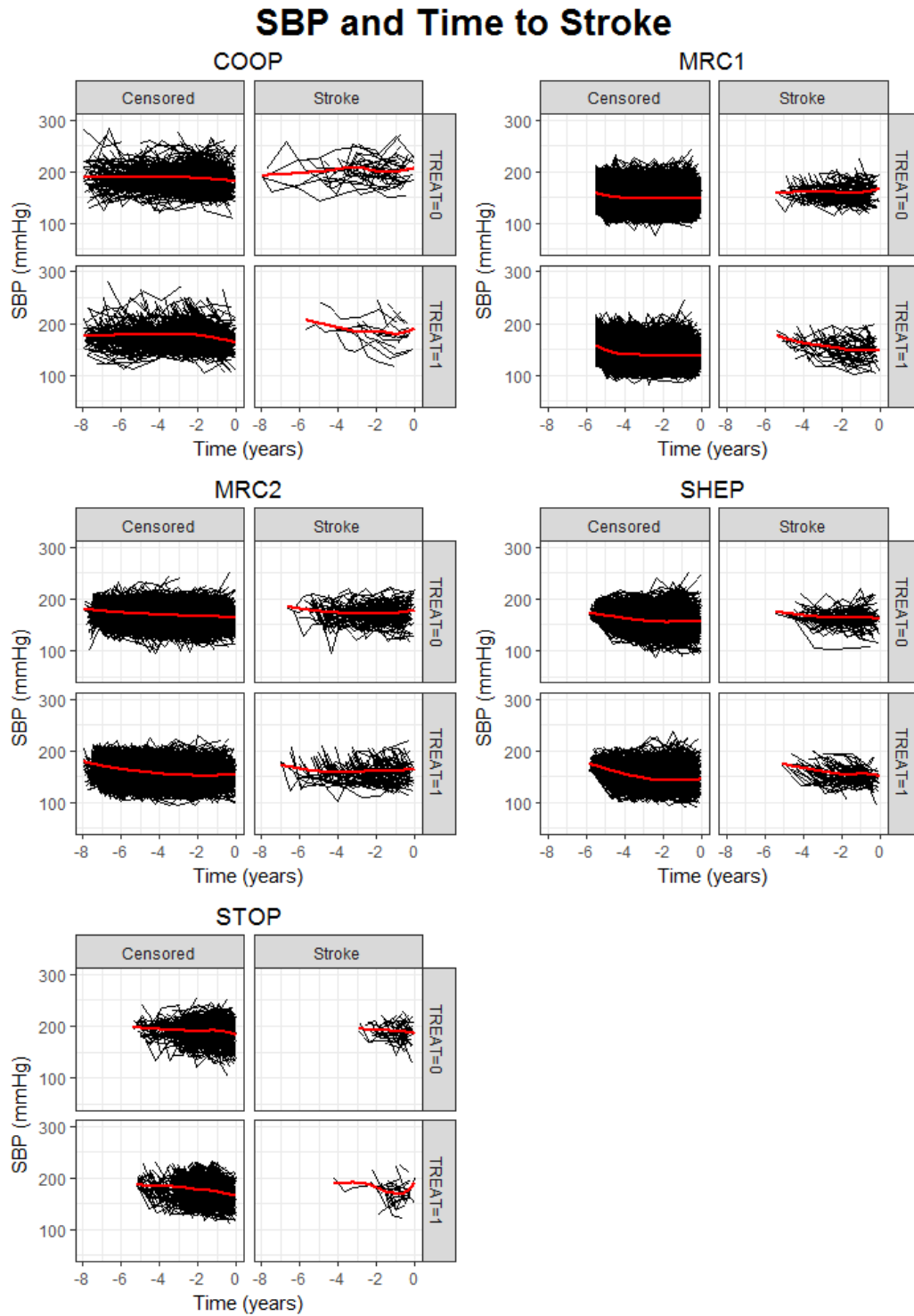


Figure 9: For SBP and time to stroke data, individual longitudinal trajectories (black) panelled by event type and treatment group (1=any drug intervention for hypertension, 0= no treatment, placebo or usual care) for each study, with loess smoother (red). Time is adjusted by subtracting individual specific survival times ( $T_{Ski}$ ) from their longitudinal measurement times ( $t_{kij}$ ).



### 6.1.3 Time-to-event data

The event rates for each outcome of death, stroke and MI in the included studies are given in Table 10. The percentage of participants experiencing death varies considerably between the included studies, with the lowest occurring in the MRC1 study (2.9%) and the highest in the EWPHE dataset (33.8%). However, the percentages of participants experiencing the other time-to-event outcomes (stroke and MI) seem more comparable across studies.

Study	Number experiencing death event (%)	Number experiencing myocardial infarction event (%)	Number experiencing stroke event (%)
<b>COOP</b>	130 (14.7)	73 (8.3)	59 (6.7)
<b>EWPHE</b>	284 (33.8)	NR	NR
<b>MRC1</b>	501 (2.9)	456 (2.6)	169 (1.0)
<b>MRC2</b>	616 (14.0)	287 (6.5)	235 (5.3)
<b>SHEP</b>	455 (9.6)	245 (5.2)	262 (5.5)
<b>STOP</b>	99 (6.1)	66 (4.1)	83 (5.1)
<b>Overall</b>	2085 (7.0)	1127 (3.8)	808 (2.7)

Table 10: Event rates for included studies in INDANA dataset. Missing data: 840 missing stroke event indicators (all from EWPHE), 848 missing myocardial infarction event indicators (840 missing from EWPHE, 8 missing from SHEP). Note that no information was recorded (NR = Not Recorded) for time to MI or time to stroke for study EWPHE.

Kaplan-Meier plots with number at risk tables are provided for the time-to-event outcomes death, MI, and stroke in Figure 10 to Figure 12. Examination of these plots identifies several points for discussion.

Firstly, for the outcome time to death (Figure 10), for studies COOP, MRC1, MRC2, and SHEP, there is little difference in the survival curves for those assigned to no treatment, placebo or usual care (red) to those assigned to any drug intervention for hypertension (blue). However, for studies EWPHE and STOP there is a greater difference between the curves. For EWPHE the curve for those assigned to placebo, no treatment or usual care is initially below that of those assigned to any drug intervention, although the curves cross at later time points. Similarly, the curve for those assigned to no treatment, placebo or usual care in the STOP trial lies below that of those assigned to any drug intervention, again

indicating worse survival for those assigned to no treatment, placebo or usual care versus any drug intervention for hypertension.

For the outcome time to MI (Figure 11), the survival curves across all studies for those assigned to any drug intervention for hypertension (blue) versus those assigned to no treatment, placebo or usual care (red) are similar. The curves are slightly more distinct for the COOP trial, especially at later time points, however still could not be described as separated from each other.

For the outcome time to stroke (Figure 12) a clearer distinction between the survival curves for those assigned to no treatment, placebo or usual care (red) versus those assigned to any drug intervention (blue) is seen. Apart from the MRC1 trial (where the survival curves are not distinct), the remaining studies all display survival curves where the curve for those assigned to no treatment, placebo or usual care falls below the curve for those assigned to treatment (indicating better survival in the any drug intervention group). In some studies (MRC2, COOP, to a point SHEP) the separation of the curves is more noticeable only after a certain time point.

Overall, for the three time-to-event outcomes, the assumption of proportional hazards appears acceptable, and time-to-event sub-model of any joint models fitted as well as any separate time-to-event analyses will follow a PH model. In the future, alternatives to the PH model will be included in the *joineRmeta* package, allowing more flexibility in model choice (see Section 8.3.2.2).

## SBP and Time to Death

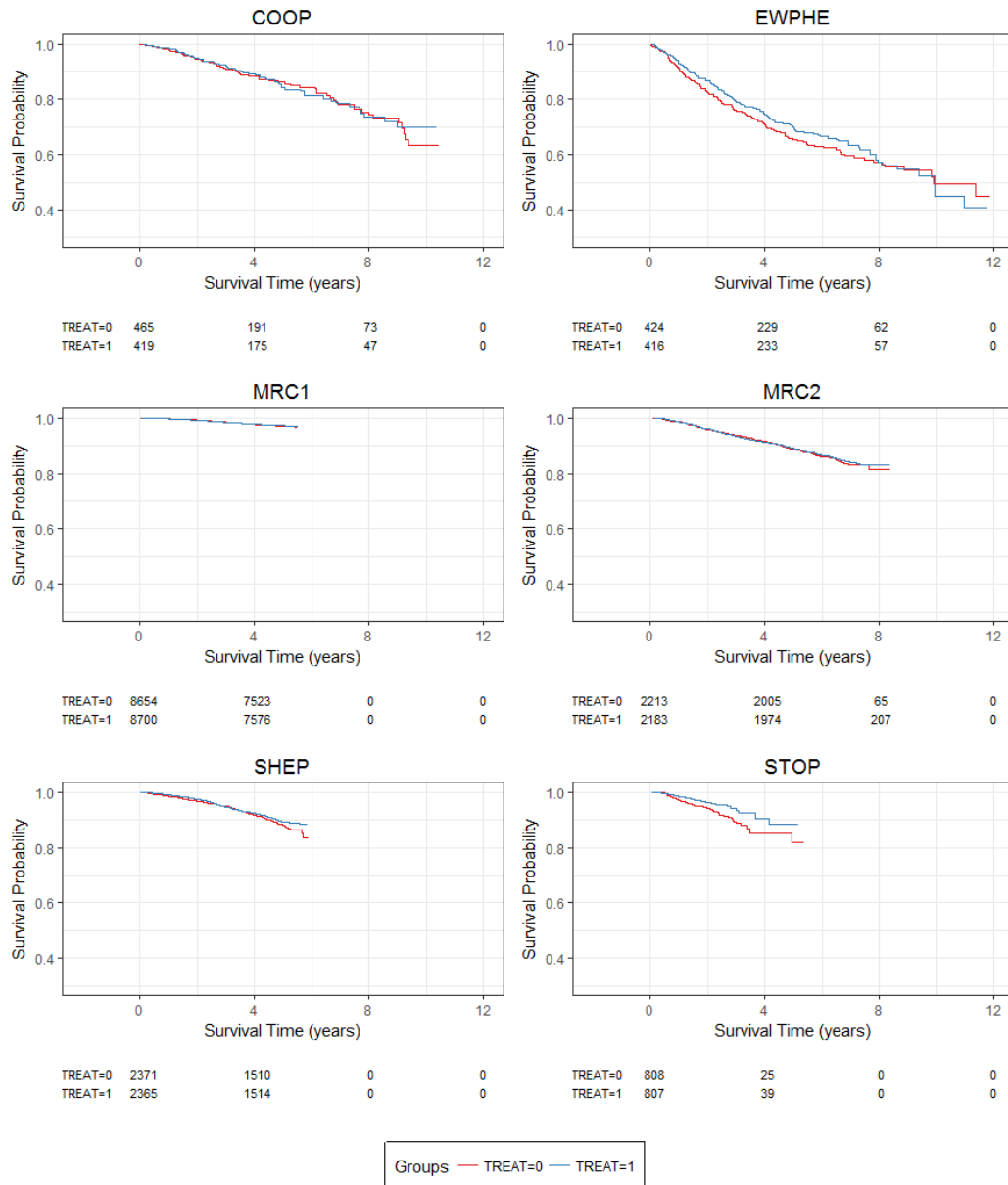


Figure 10: Kaplan-Meier curves for time to death outcome for analysis of SBP and time to death. Number at risk tables are shown under the plot for each study. The any drug intervention for hypertension group is represented by TREAT = 1, whilst the no treatment, placebo or usual care is represented by TREAT=0.

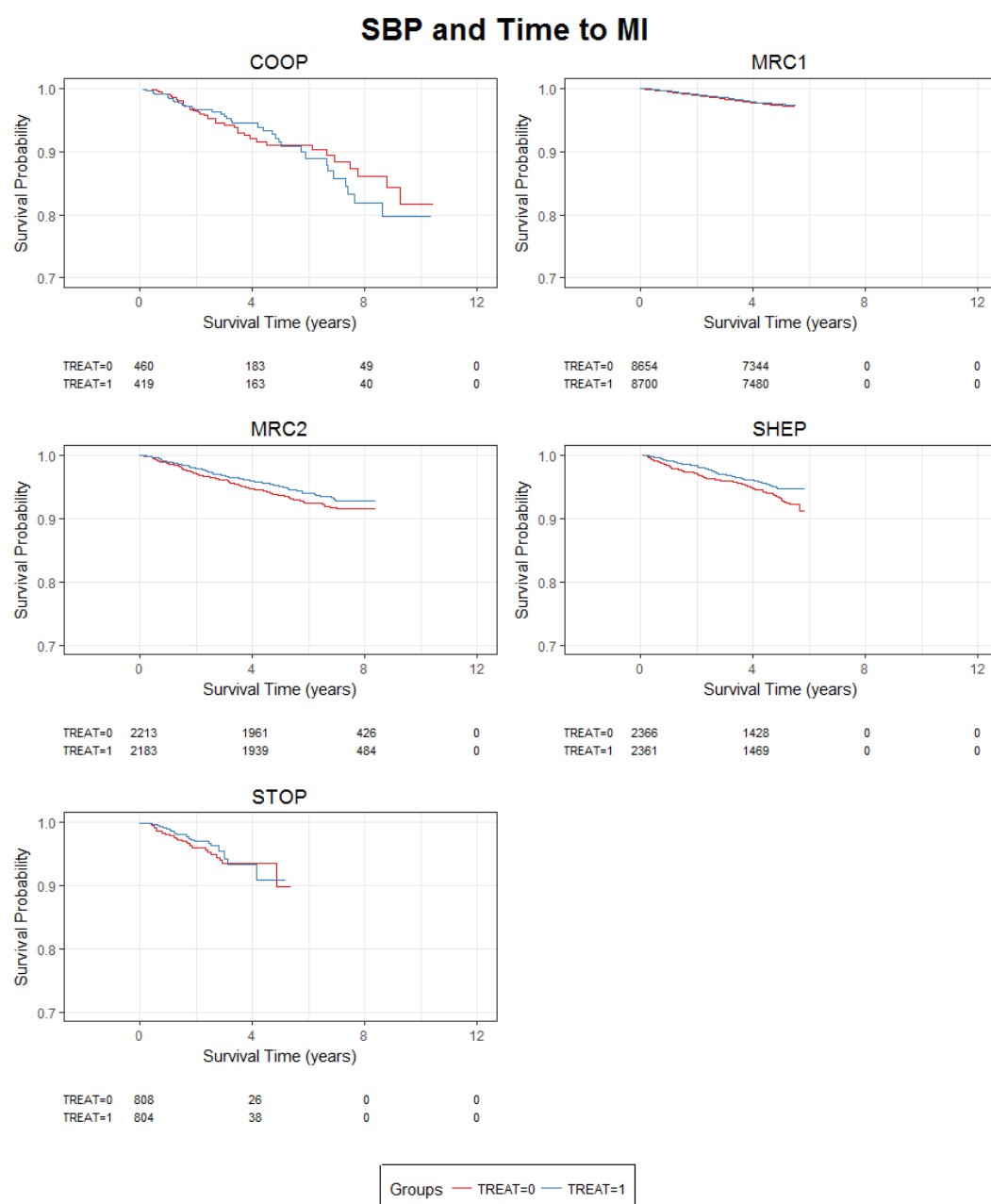


Figure 11: Kaplan-Meier curves for time to MI outcome for analysis of SBP and time to MI. Number at risk tables are shown under the plot for each study. The any drug intervention for hypertension group is represented by TREAT = 1, whilst the no treatment, placebo or usual care is represented by TREAT=0.

## SBP and Time to Stroke

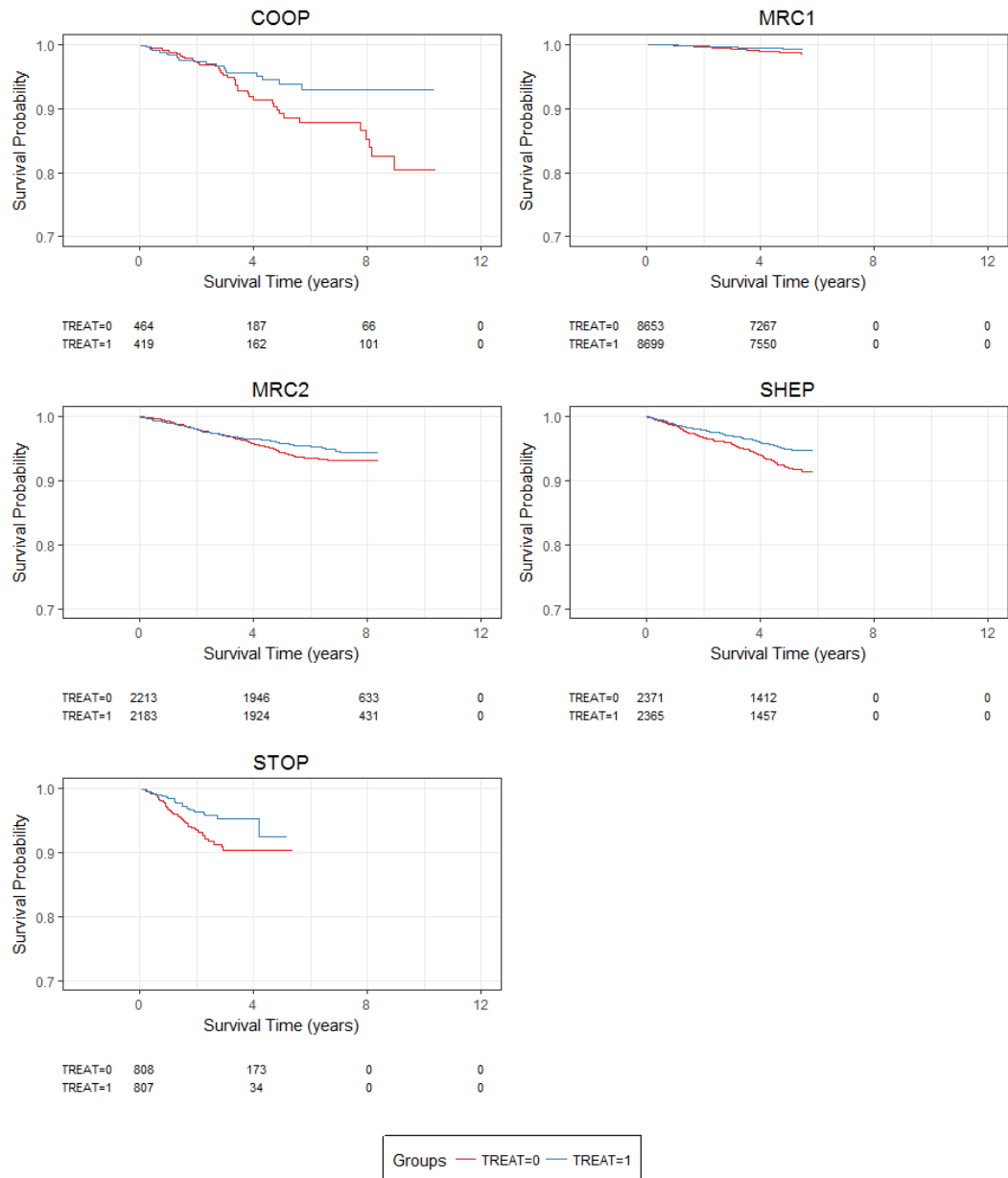


Figure 12: Kaplan-Meier curves for time to stroke outcome for analysis of SBP and time to stroke. Number at risk tables are shown under the plot for each study. The any drug intervention for hypertension group is represented by TREAT = 1, whilst the no treatment, placebo or usual care is represented by TREAT=0.

## 6.2 Two-stage Meta-Analysis of INDANA dataset

This two-stage meta-analytic analysis of the INDANA dataset was presented as a real data application in Sudell et al [176].

As discussed in Chapter 1, in single study cases, where an association exists between a longitudinal and a time-to-event outcome, joint models can provide more efficient, less biased results than separate analyses [158]. To determine whether this behaviour persists in a multi-study analysis, the investigation aimed to examine for a real dataset whether use of joint longitudinal and time-to-event models in the first stage of the MA was preferable to use of separate longitudinal or time-to-event models.

### 6.2.1 Methods

The methods and guidelines proposed in Section 3.2 for the two-stage MA of joint data are applied to the INDANA dataset. Throughout, study specific coefficient estimates include a  $k$  in their subscript e.g.  $\hat{\beta}_{12k}$  whilst the pooled estimates produced by the meta-analyses are denoted e.g.  $\hat{\beta}_{12}$ .

#### 6.2.1.1 Preliminary work

Plots of the longitudinal trajectories for each study panelled by event type were presented in Figure 4 to Figure 6 for those individuals contributing to the analysis of SBP and time to death, time to MI and time to stroke respectively. From these longitudinal trajectory plots several points were noted.

Firstly (using Figure 37 to Figure 39 in Appendix 2, which do not adjust time by survival time), the longitudinal trajectories showed a change in trajectory at approximately 6 months in each study across all time-to-event outcome groups (identifiable from the smoother applied to the trajectory plots). As discussed earlier this behaviour is accounted for through inclusion of an  $\exp(-3 * t)$  term in the longitudinal sub-model. Otherwise, it appears that the trajectories might be reasonably represented through a linear mixed effects model.

Secondly, there appears to be considerable variation between individuals in their longitudinal intercept, and the slope shown in the longitudinal trajectories. As such, a longitudinal sub-model that allows for both a random intercept and a random slope may be preferred.

Thirdly, the potential link or association between the longitudinal and the time-to-event outcomes needs to be considered. There was some evidence in Figure 4 to Figure 9 to

suggest associations between SBP and the time-to-event outcomes. The NICE guidelines “Hypertension in adults: diagnosis and management” [229] identifies firstly that for each 2 mmHg rise in SBP, there is an associated 7% increased risk of mortality from ischaemic heart disease, and 10% increased risk of mortality from stroke, and secondly that hypertension is a major risk for haemorrhagic stroke, myocardial infarction, and premature death. As such, clinical evidence exists for a link between SBP and each of time to death, time to myocardial infarction (MI) and time to stroke. As this investigation aims to investigate methods rather than change clinical practice, the random effects only proportional association structure is employed, although other association structures may have relevant clinical interpretations.

Kaplan-Meier curves plotted by treatment group (any drug intervention in blue, and no treatment, placebo or usual care in red) are given in Figure 10-Figure 12. As noted earlier (Section 6.1.3), the curves for each of time to death, time to MI and time to stroke indicate that a PH model might be generally acceptable.

From examination of both the plots, and consideration of the clinical background of the data, it is considered appropriate to analyse the data using a joint model consisting of a mixed effects model for the longitudinal sub-model, a proportional hazards model for the time-to-event sub-model, and a random effects only proportional association structure.

#### 6.2.1.2 First Stage

In the first stage of the two-stage MA of the INDANA dataset, joint models and separate longitudinal and time-to-event models, were fitted to each study. As the same joint model specification was believed appropriate across the included studies, the joint models fitted to each study  $k$  all had the following format:

$$\begin{aligned} SBP_{kij} &= \beta_{10k} + \beta_{11k}t_{kij} + \beta_{12k}treat_{ki} + \beta_{13k}\exp(-3 * t_{kij}) + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij} \quad (66) \\ \lambda_{ki}(t) &= \lambda_0(t)\exp(\beta_{21k}treat_{ki} + W_{2ki}(t)) \\ W_{2ki}(t) &= \alpha_k^{(2)}W_{1ki}(t) = \alpha_k^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) \end{aligned}$$

Here,  $t_{kij}$  is longitudinal time, and terms  $\beta_{10k}$  through  $\beta_{13k}$  are the coefficients for longitudinal fixed effects estimated for each study  $k$ . The individual level random effects  $b_{0ki}^{(2)}$  and  $b_{1ki}^{(2)}$  follow  $\mathbf{b}_{ki}^{(2)} \sim N(0, \mathbf{D}_k)$ , while the error term  $\varepsilon_{kij}$  is distributed  $N(0, \sigma_{ek}^2)$ . The time-to-event sub-model consists of an unspecified baseline hazard  $\lambda_0(t)$ , as well as a study specific fixed treatment effect with associated coefficient  $\beta_{21k}$ .

The association structure takes the zero mean individual level random effects from the longitudinal sub-model and inserts them with study specific common coefficient  $\alpha_k^{(2)}$  (the association parameter) into the time-to-event sub-model giving term  $W_{2ki}(t)$ .

As noted earlier, as well as the joint models, standalone longitudinal and time-to-event models were fitted to the INDANA datasets. These standalone models had the same specifications as the corresponding sub-models of the joint model shown in equation (66), apart from that the  $W_{2ki}(t)$  term was not present in the standalone time-to-event model.

#### 6.2.1.3 Second stage

The second stage of the MA involved extracting parameters of interest from the study specific model fits and pooling them using standard meta-analytic techniques. During the second stage of the meta-analysis, interest focussed on the fixed treatment assignment coefficient from both sub-models (namely coefficients  $\beta_{12}$  and  $\beta_{21}$ ) as well as the association parameter  $\alpha^{(2)}$ . These coefficient estimates were extracted from the models for each study for each analysis, as well as the standard error estimate and the sample size.

These results were then pooled in both a fixed and a random effects MA, using the inverse variance approach [2, 184], with the DerSimonian and Laird [10] approach for the random MA (methods discussed in Section 3.2.3). All studies with available data contribute to the analysis. The same model specification was utilised for each study, for each outcome pair (SBP and time to death, SBP and time to MI, and SBP and time to stroke). Separate MA were performed for each parameter of interest (namely the longitudinal treatment effect estimate  $\hat{\beta}_{12}$ , the time-to-event treatment effect estimate  $\hat{\beta}_{21}$ , and the association parameter estimate  $\hat{\alpha}^{(2)}$ ).

#### 6.2.1.4 Software

Joint models, and separate longitudinal and time-to-event models were fitted using functions in the R package *joineRmeta* (available through GitHub: <https://github.com/mesudell/joineRmeta>). The functions available in this packages were developed during this thesis, and are fully described in Chapter 4. During the second stage of the two-stage MA, the R package *meta* [182] was used.

To obtain standard errors in the *joineRmeta* package it is necessary to perform a bootstrapping procedure; Hsieh et al [89] emphasised that use of the profile likelihood to estimate the standard errors in joint models could lead to underestimation of the standard errors when an unspecified baseline hazard is employed. Consequently the time taken to



obtain precision estimates from the joint models was considerably longer than for the separate meta-analyses given some of the larger sample sizes in the included studies. This is further discussed in the simulation studies in Chapter 7.

### 6.2.2 Results

During the two-stage MA, a statistically significant negative time-to-event treatment assignment coefficient ( $\beta_{21}$ ) would indicate that assignment to any drug intervention for hypertension versus no treatment, placebo or usual care significantly reduced the risk of the event in question. A statistically significant negative longitudinal treatment assignment coefficient ( $\beta_{12}$ ) would indicate that assignment to any drug intervention for hypertension significantly decreased SBP. A statistically significant positive association parameter ( $\alpha^{(2)}$ ) would indicate that individuals with a positive deviation above the population mean longitudinal value in their recorded longitudinal values at a given time point are at higher risk of the event.

Table 11 to Table 13 present the results of the analysis of the INDANA dataset in a two-stage MA employing separate longitudinal models, separate time-to-event models or joint models in the first stage of the analysis, for the investigation of SBP and each of time to death (Table 11), time to MI (Table 12) and time to stroke (Table 13).

#### 6.2.2.1 Longitudinal component

There is little difference between the estimates of the longitudinal treatment effect coefficient from the separate longitudinal analysis compared to the joint analysis (Table 11 to Table 13). Assignment to treatment for hypertension versus no treatment, placebo or usual care is estimated to significantly reduce SBP for each analysis conducted. There is heterogeneity between study specific estimates (the  $\tau^2$  statistic varied between 2.92 and 3.21, whilst the  $I^2$  statistic was greater than 95% in all cases). However, both the fixed and random effects MA agreed in the significance and direction of the pooled results.

#### 6.2.2.2 Time-to-event component

Again, there was similarity for this real data example between the time-to-event treatment effect coefficient estimated within each study between the separate and the joint model analyses for each set of outcomes (Table 11 to Table 13). Additionally, there was little evidence of heterogeneity between studies across any of the meta-analyses for either method for the time-to-event treatment coefficient for any of time to death, MI or stroke.

For time to death (Table 11), for both the separate and the joint analysis, a significant negative time-to-event treatment effect coefficient was observed in the STOP trial, however no significant effect of treatment assignment was observed in any other study for SBP and time to death, or in the pooled results from either the fixed or random MA.

For time to MI (Table 12), a significant negative time-to-event treatment effect (indicating assignment to any drug intervention for hypertension significantly reducing the risk of MI) was observed for both the separate and the joint analysis in the SHEP trial. However, again, none of the remaining trials observed a significant treatment effect. Interestingly, the pooled results from both the fixed and the random effects MA for the separate time-to-event analysis showed a significant negative pooled treatment effect coefficient. However, neither of the pooled results from the joint analysis for the fixed or the random MA were significant. As such, simultaneously modelling both the longitudinal and the time-to-event outcome resulted in a shift in the estimated pooled treatment effect coefficient for the time-to-event outcome towards zero.

For time to stroke (Table 13), all studies for the separate time-to-event analysis, and all studies apart from MRC2 for the joint analysis, reported a significant negative treatment effect coefficient (meaning that most cases observed that assignment to treatment was linked to a significant reduction in risk of stroke). The pooled results for both the separate and the joint analyses were similar, with both the fixed and random MA displaying significant negative pooled treatment effect coefficients. This indicated that both separate and joint methods estimated that assignment to any hypertensive treatment version no treatment, placebo or usual care decreased the risk of stroke.

#### 6.2.2.3 Association

The estimated association parameters from each study for each outcome combination (SBP and each of time to death, time to MI and time to stroke, Table 11 to Table 13) were small in magnitude. Additionally, for each combination of outcomes, there was evidence of heterogeneity between studies in the estimated association parameters ( $I^2$  ranging from 83.20% to 90.30%, p value for chi squared test for heterogeneity <0.001 in all cases). In all three cases  $\tau^2$  was small in magnitude, however the magnitudes of the association parameters were in themselves small. For SBP and time to death, and SBP and time to MI (Table 11 and Table 12) a significant positive association parameter was estimated in the largest included trial, MRC1, suggesting that an individual with SBP values above the population average at a given time has an associated higher risk of an event at that time.

The pooled association parameter estimate was significant for both these analyses for the fixed MA, however in both cases the pooled estimate from the random MA was not significant. For time to stroke (Table 13), a significant positive association parameter was estimated for studies COOP, MRC1 and SHEP. Again, the fixed MA produced a significant positive pooled association parameter, however the result from the random MA was not significant. Throughout, the wider confidence intervals produced by the random MA compared to the fixed MA were expected given the high level of observed heterogeneity.

Study	N (no. of events) [no. longitudinal measurements]	Longitudinal: Treatment coefficient $\beta_{12}$ (95% CI)		Time-to-event: Treatment coefficient $\beta_{21}$ (95% CI)		Association Parameter $\alpha^{(2)}$ (95% CI)
		Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	Joint Sub-Model Results
Time to death						
Estimates from each study						
COOP	884 (130) [4621]	-10.9 (-13.11, -8.69)	-10.9 (-13.3, -8.54)	-0.04 (-0.38, 0.31)	-0.05 (-0.46, 0.32)	-0.003 (-0.021, 0.012)
EWPHE	840 (284) [3825]	-11.38 (-13.44, -9.32)	-11.38 (-13.78, -9.09)	-0.10 (-0.33, 0.13)	-0.06 (-0.32, 0.18)	0.007 (-0.004, 0.018)
MRC1	17354 (501) [101078]	-8.18 (-8.57, -7.78)	-8.18 (-8.52, -7.81)	-0.03 (-0.20, 0.15)	0.02 (-0.18, 0.18)	<b>0.034 (0.024, 0.042)</b>
MRC2	4396 (616) [26457]	-10.66 (-11.31, -10.01)	-10.66 (-11.46, -10.07)	-0.04 (-0.19, 0.12)	-0.04 (-0.18, 0.11)	-0.001 (-0.013, 0.011)
SHEP	4736 (455) [20816]	-6.99 (-7.59, -6.38)	-6.99 (-7.64, -6.40)	-0.13 (-0.32, 0.05)	-0.11 (-0.31, 0.07)	0.008 (-0.001, 0.016)
STOP	1615 (96) [5777]	-11.94 (-13.27, -10.62)	-11.94 (-13.48, -10.43)	<b>-0.51 (-0.92, -0.10)</b>	<b>-0.56 (-1.12, -0.15)</b>	-0.011 (-0.04, 0.012)
Meta-analysis						
Fixed MA	29825 (2082) [162574]	<b>-8.66 (-8.94, -8.38)</b>	<b>-8.62 (-8.9, -8.34)</b>	-0.08 (-0.17, 0.00)	-0.06 (-0.15, 0.03)	<b>0.01 (0.01, 0.02)</b>
Random MA	29825 (2082) [162574]	<b>-9.86 (-11.4, -8.33)</b>	<b>-9.83 (-11.34, -8.32)</b>	-0.09 (-0.17, 0.00)	-0.06 (-0.16, 0.03)	0.01 (-0.01, 0.02)
$\tau^2$		3.2091	3.0541	6.00E-04	0.0015	2.00E-04
$I^2$		95.20%	94.90%	4.30%	10.90%	83.80%
p value		< 0.0001	< 0.0001	0.3892	0.3456	< 0.0001

Table 11: Results for two-stage analysis of SBP and time to death from the INDANA data ( $\beta_{12}$ ,  $\beta_{21}$  and  $\alpha^{(2)}$  are as defined in equation (66), p value is for chi squared test for presence of significant heterogeneity).

Study	N (no. of events) [no. longitudinal measurements]	Longitudinal: Treatment coefficient $\beta_{12}$ (95% CI)		Time-to-event: Treatment coefficient $\beta_{21}$ (95% CI)		Association Parameter $\alpha^{(2)}$ (95% CI)
		Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	
Time to MI						
Estimates from each study						
COOP	884 (73) [4573]	-11.1 (-13.32, -8.89)	-11.10 (-13.19, -8.70)	0.03 (-0.43, 0.49)	0.03 (-0.46, 0.40)	-0.001 (-0.019, 0.013)
MRC1	17354 (456) [100699]	-8.18 (-8.58, -7.79)	-8.18 (-8.61, -7.82)	-0.06 (-0.24, 0.12)	-0.02 (-0.19, 0.17)	<b>0.032 (0.024, 0.041)</b>
MRC2	4396 (287) [26303]	-10.70 (-11.36, -10.05)	-10.70 (-11.46, -10.14)	-0.21 (-0.44, 0.02)	-0.18 (-0.39, 0.04)	0.018 (0.000, 0.035)
SHEP	4728 (245) [20609]	-7.01 (-7.61, -6.40)	-7.01 (-7.63, -6.39)	<b>-0.32 (-0.57, -0.06)</b>	<b>-0.31 (-0.62, -0.09)</b>	0.003 (-0.011, 0.015)
STOP	1615 (63) [5739]	-11.96 (-13.28, -10.64)	-11.96 (-13.38, -10.52)	-0.23 (-0.72, 0.27)	-0.24 (-0.84, 0.26)	-0.004 (-0.035, 0.018)
Meta-analysis						
Fixed MA	28977 (1124) [157923]	-8.62 (-8.91, -8.34)	-8.59 (-8.87, -8.30)	<b>-0.16 (-0.27, -0.04)</b>	-0.12 (-0.23, 0.00)	<b>0.02 (0.01, 0.02)</b>
Random MA	28977 (1124) [157923]	-9.66 (-11.32, -8.01)	-9.64 (-11.25, -8.02)	<b>-0.16 (-0.27, -0.04)</b>	-0.12 (-0.23, 0.00)	0.01 (0.00, 0.03)
$\tau^2$		3.2083	3.0179	0	0	3.00E-04
$I^2$		96%	95.50%	0%	0.10%	83.20%
p value		< 0.0001	< 0.0001	0.4754	0.4053	< 0.0001

Table 12: Results for two-stage analysis of SBP and time to MI from the INDANA data ( $\beta_{12}$ ,  $\beta_{21}$  and  $\alpha^{(2)}$  are as defined in equation (66), p value is for chi squared test for presence of significant heterogeneity)

Study	N (no. of events) [no. longitudinal measurements]	Longitudinal: Treatment coefficient $\beta_{12}$ (95% CI)		Time-to-event: Treatment coefficient $\beta_{21}$ (95% CI)		Association Parameter $\alpha^{(2)}$ (95% CI)
		Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	Joint Sub-Model Results
Time to stroke						
Estimates from each study						
COOP	884 (59) [4527]	-10.73 (-12.94, -8.51)	-10.73 (-12.96, -8.33)	-0.58 (-1.12, -0.04)	-0.48 (-1.23, -0.02)	0.034 (0.016, 0.053)
MRC1	17354 (169) [100914]	-8.18 (-8.58, -7.79)	-8.18 (-8.59, -7.81)	-0.60 (-0.92, -0.29)	-0.51 (-0.87, -0.24)	0.061 (0.048, 0.072)
MRC2	4396 (235) [26251]	-10.66 (-11.31, -10.00)	-10.66 (-11.29, -9.98)	-0.28 (-0.54, -0.02)	-0.24 (-0.49, 0.00)	0.02 (-0.003, 0.039)
SHEP	4736 (262) [20441]	-6.96 (-7.56, -6.35)	-6.96 (-7.59, -6.45)	-0.45 (-0.70, -0.20)	-0.39 (-0.62, -0.15)	0.021 (0.010, 0.030)
STOP	1615 (83) [5701]	-11.70 (-13.03, -10.37)	-11.70 (-13.29, -10.15)	-0.64 (-1.10, -0.19)	-0.70 (-1.21, -0.24)	-0.014 (-0.038, 0.01)
Meta-analysis						
Fixed MA	28985 (808) [157834]	-8.59 (-8.87, -8.30)	-8.56 (-8.84, -8.28)	-0.46 (-0.60, -0.32)	-0.39 (-0.53, -0.25)	0.03 (0.02, 0.04)
Random MA	28985 (808) [157834]	-9.53 (-11.14, -7.92)	-9.5 (-11.09, -7.91)	-0.46 (-0.60, -0.32)	-0.39 (-0.53, -0.25)	0.03 (0.00, 0.05)
$\tau^2$		3.0349	2.9163	0	0	6.00E-04
$I^2$		95.80%	95.70%	0%	0%	90.30%
p value		< 0.0001	< 0.0001	0.4694	0.4666	< 0.0001

Table 13: Results for two-stage analysis of SBP and time to stroke from the INDANA data ( $\beta_{12}$ ,  $\beta_{21}$  and  $\alpha^{(2)}$  are as defined in equation (66), p value is for chi squared test for presence of significant heterogeneity)

### 6.3 One-stage Meta-Analysis of INDANA dataset

In this section the one-stage meta-analysis of the INDANA dataset is presented. This section commences with statements of the exact models fitted to the data, followed by presentation of the results of the one-stage model.

#### 6.3.1 Methods

##### 6.3.1.1 *Model specifications*

As in the two-stage analysis (Section 6.2), each pairwise combination of longitudinal outcome SBP and time to death, time to MI, and time to stroke have been analysed. The models discussed in Section 3.3 have been employed, which take format as in Table 14 when applied to this dataset. Models are grouped to demonstrate different methods to account for between study heterogeneity. Group 0 ignores between study differences and Group 1 includes fixed interaction terms (in both sub-models) between the study membership variable and treatment assignment. Groups 2 and 3 introduce study level random effects (with random intercept, or random intercept and treatment term respectively). Groups 4 and 5 stratify the baseline hazard of the time-to-event sub-model by study, with Group 5 additionally containing a study level random treatment term. Again, the  $\exp(-3 * t_{kij})$  term is present in the longitudinal sub-model, to model the change in trajectory noted earlier.

Once models were fitted, estimates of the coefficients of interest were extracted from the models (treatment effects  $\beta_{12}$  and  $\beta_{21}$  from each sub-model, and any association parameters available) along with their 95% confidence intervals, and the number of individuals contributing to the analyses.

Model Group	Model component	Equation
0	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
1	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + \beta_{14}study_{ki} + \beta_{15}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + \beta_{23}treat_{ki} * study_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
2	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + \beta_{14}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$
3	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki})$
4	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + \beta_{14}study_{ki} + \beta_{15}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
5	Longitudinal Sub-Model	$SBP_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13} \exp(-3 * t_{kij}) + \beta_{14}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}time_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$

Table 14: Specification of one-stage models fitted to INDANA dataset



### 6.3.1.2 *Software*

Models were fitted using the *joinerMeta* package (described in Chapter 4 and Chapter 5). Joint models in a single study setting can be time-consuming to fit. In the models examined here this issue persists through the need to bootstrap to gain standard error estimates. Both model fitting, and so bootstrapping, was time intensive due to the large size of the meta-dataset. As such, the Chadwick supercomputer (owned by the University of Liverpool), was used to fit the models, and run each separate bootstrap, which were then compiled on a standalone laptop. One-stage methods may be difficult for researchers to apply without access to such resources, dependent on the size of the meta-dataset.

### 6.3.2 *Results*

The results for the one-stage analysis of the INDANA dataset are shown in Table 15-Table 17 and Figure 13-Figure 21. Results are displayed for each of the six model groups (Table 14) for both the joint and separate analyses for each pair of outcomes. As noted earlier, the separate models have the same structure (excepting association terms) as the corresponding sub-model of the joint model.

#### 6.3.2.1 *SBP and time to death*

For the analysis of SBP and time to death, across all groups of models fitted for both the separate and joint analyses, all estimates of longitudinal treatment effect were significant and negative. This indicated that allocation to any drug intervention for hypertension versus no treatment, placebo or usual care significantly reduced SBP (Table 15). All groups of joint models tested gave similar results apart from group 3, which accounted for between study heterogeneity solely through use of study level random effects. This estimate, whilst still being significant and negative, was much closer to zero than the joint model longitudinal treatment coefficient estimate from other groups. The estimates within each model group were similar between the joint and the separate methods (apart from group 3, where the separate model gave an estimate comparable to the other groups). However, the confidence intervals were wider in many cases for the separate than the joint models (Figure 13).

The majority of time-to-event treatment effect estimates for SBP and time to death (Table 15) were not statistically significant (apart from the treatment effect estimate from the STOP trial for the joint and separate models in model group 1). Again, there was similarity between the estimates between the joint and separate analyses, although the results from the joint model were closer to zero, and variation existed in the confidence for models

belonging to group 1 (Figure 14). Overall, there was little evidence from any model group, from either the separate or joint analyses indicating any difference in risk of death between the any drug intervention for hypertension group versus no treatment, placebo or usual care.

For all model groups, the estimated individual level association parameters for the SBP and time to death analysis were positive and significant, although small in magnitude (Table 15, Figure 15). Estimates were similar across all the model groups, with interpretation that individuals with higher SBP values than the population average were at higher risk of death.

The estimates of the study level association parameter (Table 15, Figure 15) were insignificant for groups 2 and 5 (model groups involving just a study level random intercept), and significant and positive for group 3 (the model group involving a study level random intercept and random treatment effect). This significant study level association parameter would be interpreted that study populations with a study average SBP higher than the overall population average are at higher risk of death. The association estimate for the group 3 model differed significantly in value from those from model groups 2 or 5, and had a narrower confidence interval. The variation in study level association parameter estimates may be due to the estimation of study level random effects being based on 6 data points – the number of studies in the analysis. Consequently, their estimation may be unreliable.

### SBP and time to Death

Model Group	Longitudinal Treatment Effect Parameter(s)		Time-to-Event Treatment Effect Parameter(s)		Association parameters
	Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	
0	$\beta_{12}$ -9.52 (-9.90, -9.13)	-9.52 (-9.92, -9.19)	$\beta_{21}$ -0.09 (-0.17, 0.00)	-0.02 (-0.13, 0.07)	$\alpha^{(2)}$ <b>0.032 (0.029, 0.035)</b>
1	$\beta_{12COOP}$ -10.03 (-11.74, -8.33)	-10.04 (-12.39, -7.91)	$\beta_{21COOP}$ -0.07 (-0.18, 0.04)	0.02 (-0.37, 0.41)	$\alpha^{(2)}$ <b>0.013 (0.009, 0.019)</b>
	$\beta_{12EWPHE}$ -13.15 (-16.56, -9.74)	-13.15 (-15.24, -11.10)	$\beta_{21EWPHE}$ -0.13 (-0.35, 0.09)	-0.03 (-0.31, 0.25)	
	$\beta_{12MRC1}$ -7.78 (-9.57, -5.99)	-7.78 (-8.17, -7.42)	$\beta_{21MRC1}$ -0.06 (-0.48, 0.37)	0.00 (-0.16, 0.15)	
	$\beta_{12MRC2}$ -10.72 (-11.10, -10.34)	-10.72 (-11.33, -10.07)	$\beta_{21MRC2}$ -0.07 (-0.51, 0.37)	-0.01 (-0.16, 0.16)	
	$\beta_{12SHEP}$ -8.30 (-9.06, -7.55)	-8.31 (-8.88, -7.75)	$\beta_{21SHEP}$ -0.16 (-0.56, 0.23)	-0.11 (-0.31, 0.09)	
	$\beta_{12STOP}$ -14.16 (-14.91, -13.40)	-14.16 (-15.40, -12.93)	$\beta_{21STOP}$ <b>-0.54 (-0.95, -0.13)</b>	<b>-0.49 (-0.95, -0.14)</b>	
2	$\beta_{12}$ -10.62 (-12.68, -8.57)	-10.63 (-11.18, -9.97)	$\beta_{21}$ -0.08 (-0.17, 0.00)	-0.05 (-0.15, 0.04)	$\alpha^{(2)}$ <b>0.013 (0.008, 0.018)</b> $\alpha^{(3)}$ 0.000 (-0.043, 0.052)
3	$\beta_{12}$ -10.67 (-12.67, -8.67)	-2.70 (-3.09, -2.42)	$\beta_{21}$ -0.09 (-0.17, 0.00)	-0.05 (-0.14, 0.03)	$\alpha^{(2)}$ <b>0.011 (0.007, 0.016)</b> $\alpha^{(3)}$ <b>0.052 (0.049, 0.055)</b>
4	$\beta_{12COOP}$ -10.03 (-11.74, -8.33)	-10.04 (-12.31, -8.07)	$\beta_{21}$ -0.08 (-0.17, 0.00)	-0.06 (-0.13, 0.03)	$\alpha^{(2)}$ <b>0.013 (0.008, 0.017)</b>
	$\beta_{12EWPHE}$ -13.15 (-16.56, -9.74)	-13.15 (-15.32, -11.19)			
	$\beta_{12MRC1}$ -7.78 (-9.57, -5.99)	-7.78 (-8.23, -7.43)			
	$\beta_{12MRC2}$ -10.72 (-11.10, -10.34)	-10.72 (-11.36, -10.11)			
	$\beta_{12SHEP}$ -8.30 (-9.06, -7.55)	-8.30 (-8.94, -7.59)			
	$\beta_{12STOP}$ -14.16 (-14.91, -13.40)	-14.15 (-15.25, -12.90)			
5	$\beta_{12}$ -10.62 (-12.68, -8.57)	-10.63 (-11.17, -10.06)	$\beta_{21}$ -0.08 (-0.17, 0.00)	-0.06 (-0.14, 0.03)	$\alpha^{(2)}$ <b>0.013 (0.006, 0.017)</b> $\alpha^{(3)}$ 0.000 (-0.039, 0.048)

Table 15: One-stage joint and separate model results for analysis of SBP and time to death by model group (dataset contains 29825 individuals, 2082 events, and 162574 longitudinal measurements)

# Longitudinal Treatment Effect

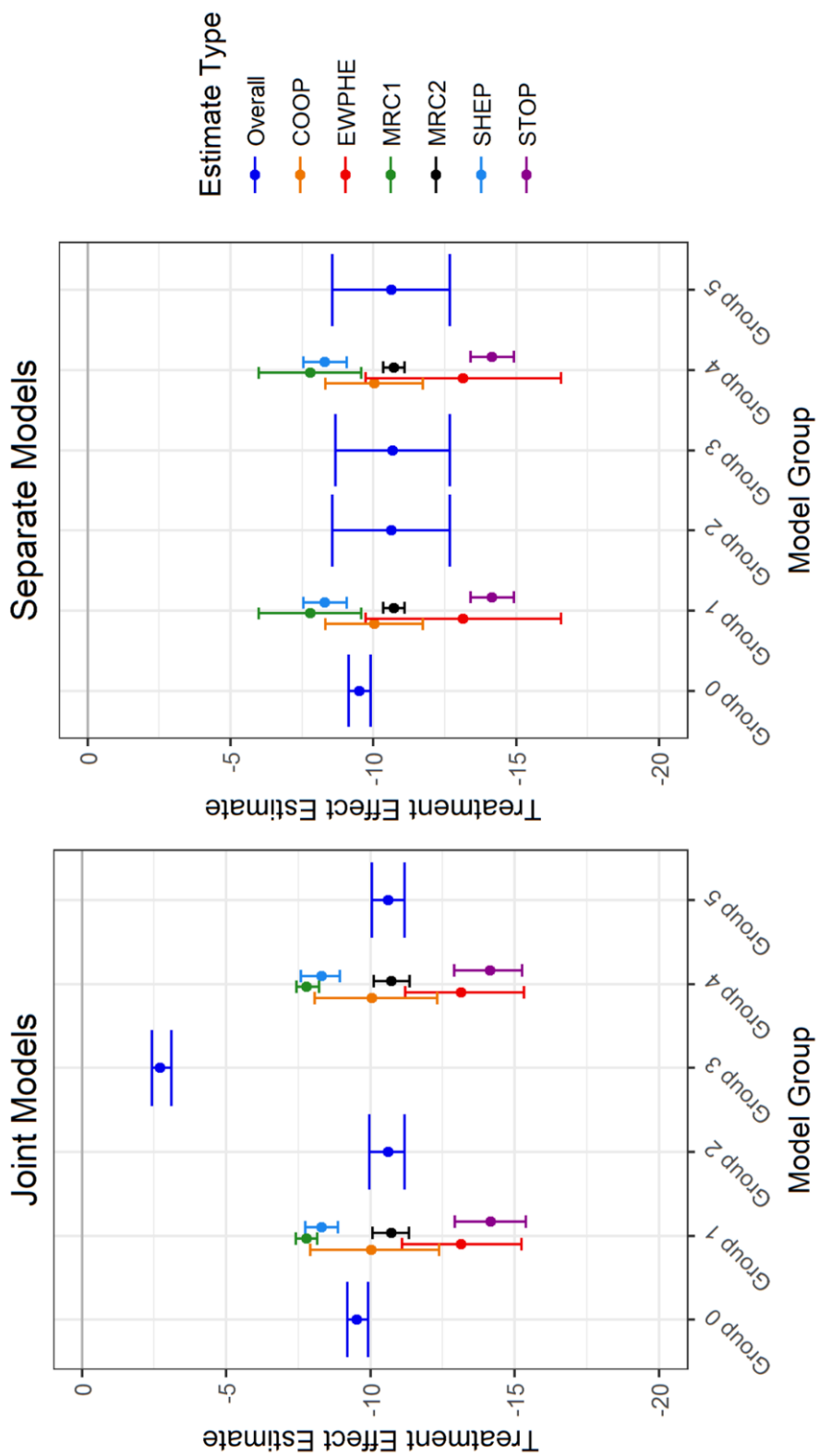


Figure 13: Graphical representation of estimates of longitudinal treatment effect shown in Table 15 for one-stage analysis of SBP and time to death

# Time-to-Event Treatment Effect

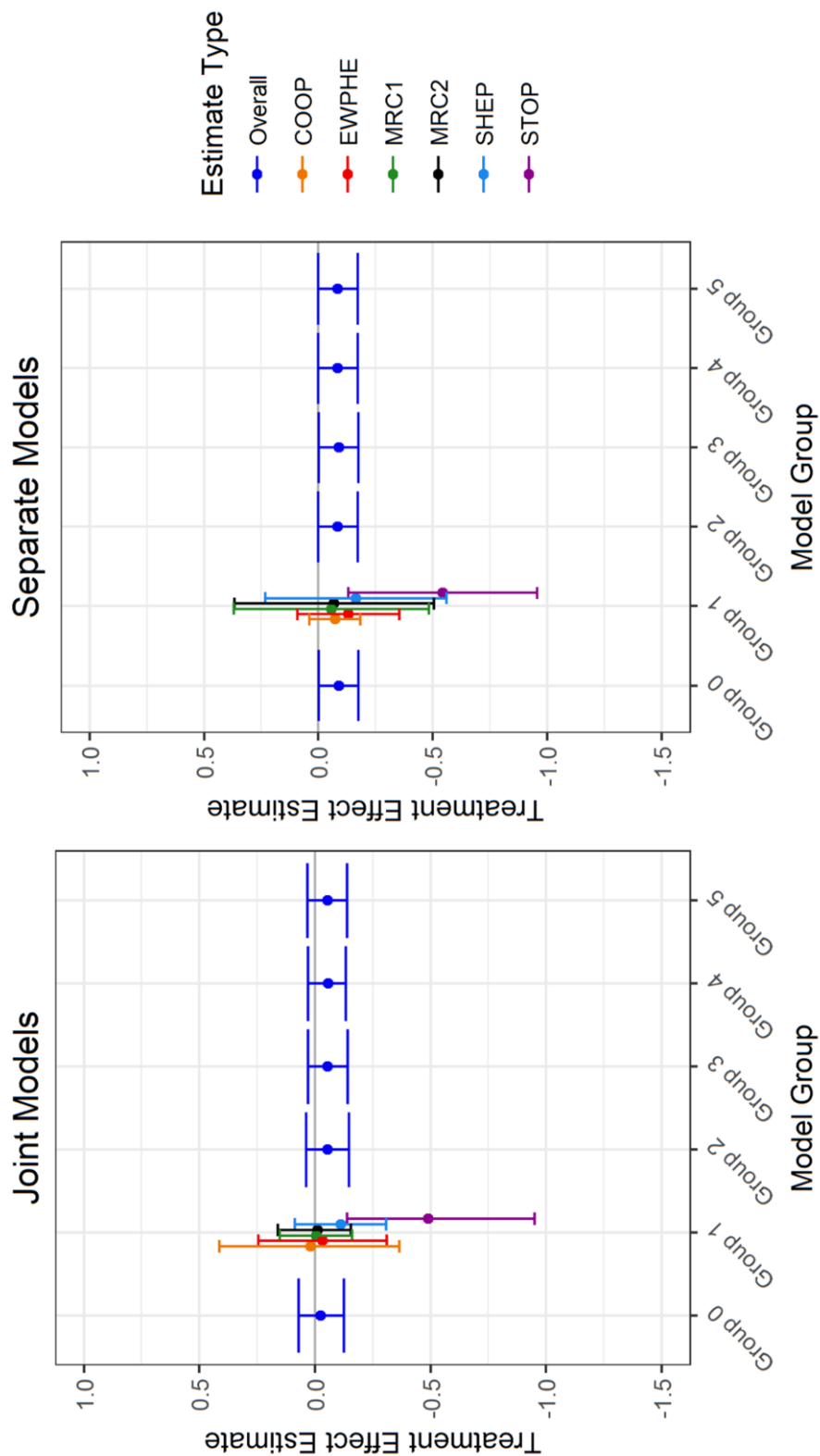


Figure 14: Graphical representation of estimates of time-to-event treatment effect shown in Table 15 for one-stage analysis of SBP and time to death

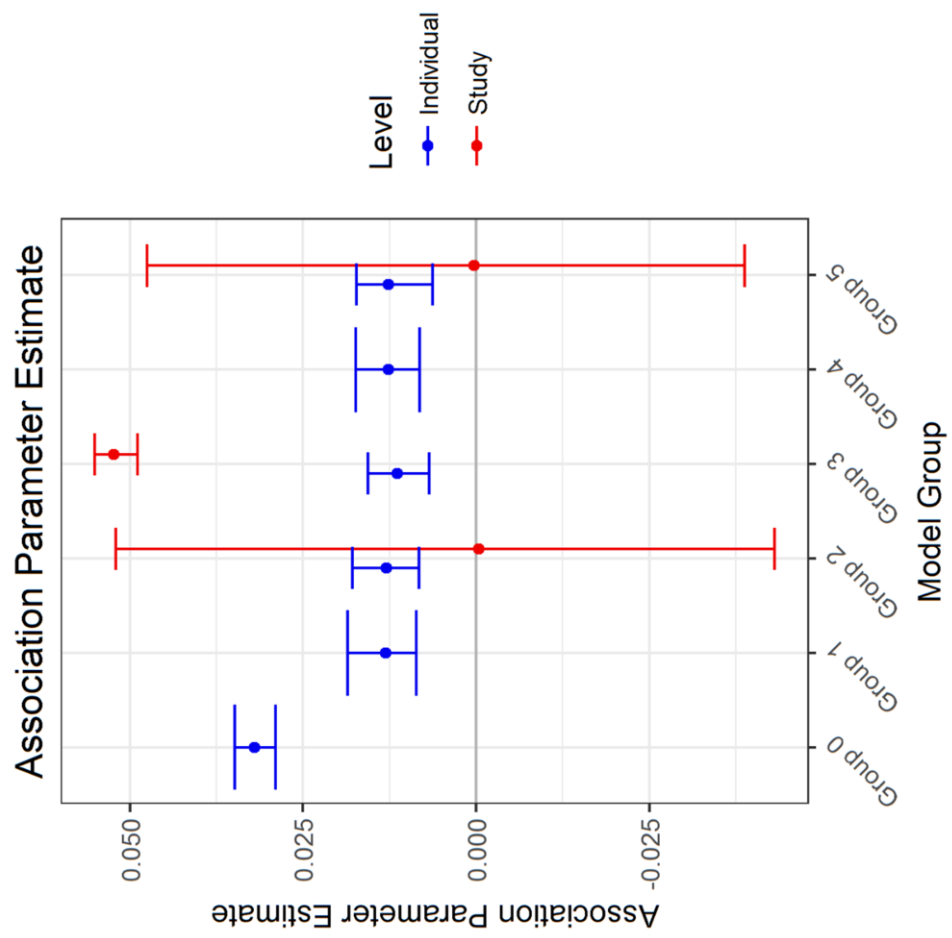


Figure 15: Graphical representation of estimates of association parameters shown in Table 15 for one-stage analysis of SPP and time to death

### 6.3.2.2 *SBP and time to MI*

For SBP and time to MI, all estimates of longitudinal treatment effect produced by either the joint model or the separate analyses from any model group were significant and negative (Table 16). As such, the analyses indicated that assignment to any drug intervention for hypertension versus no treatment, placebo or usual care significantly reduced SBP. The estimate produced by the joint model for model group 3 (which accounted for between study heterogeneity using study level random intercept and treatment terms) was different in magnitude to the other joint model estimates. Except for model group 3, treatment effect estimates from separate and joint models were comparable. However, confidence intervals (Figure 16) were wider for the separate analyses for model groups 2, 3 and 5 (which each contained study level random effects), and differed between the joint and separate analyses for groups 1 and 4 (which contained fixed between study membership and treatment group in the longitudinal sub-model).

The time-to-event treatment effect coefficient for SBP and time to MI showed evidence of a significant reduction in risk of MI for those assigned to any drug intervention for hypertension versus no treatment, placebo or usual care for the joint and separate analyses from model groups 0, 2, 3, 4, 5 (Table 16). However, for model group 1 (which included fixed study membership and interaction between study membership and treatment assignment), only the estimate from SHEP for the joint model was significant. Overall, there was agreement between the treatment effect estimates from the joint and separate analyses across the model groups (Figure 17), although the results for the joint models were closer to zero, and there was variation in the confidence intervals between the joint and separate results for group 1 models.

All estimates of the individual level association parameter for the SBP and time to MI across all model groups were significant and positive (Table 16, Figure 18), indicating that individuals with higher than average values of SBP were at greater risk of MI. However, the estimates were small in magnitude.

The study level association parameters (Table 16, Figure 19), were non-significant for the model groups that just contained a study level intercept (groups 2 and 5), however group 3 (which contained a study level random intercept and treatment effect) reported a significant positive association parameter. This would be interpreted that study populations with a study average SBP higher than the overall population average are at

higher risk of MI. However, again due to the number of studies included in the meta-analysis, study level random effects may be poorly estimated.



### SBP and time to MI

Model Group	Longitudinal Treatment Effect Parameter(s)		Time-to-Event Treatment Effect Parameter(s)		Association parameters Joint Sub-Model Results
	Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	
0	$\beta_{12}$ -9.45 (-9.84, -9.07)	-9.46 (-9.82, -8.98)	$\beta_{21}$ -0.16 (-0.28, -0.04)	-0.13 (-0.24, -0.02)	$\alpha^{(2)}$ 0.027 (0.023, 0.031)
1	$\beta_{12COOP}$ -10.18 (-11.88, -8.48)	-10.18 (-12.65, -8.08)	$\beta_{21COOP}$ -0.13 (-0.26, 0.00)	0.11 (-0.43, 0.53)	$\alpha^{(2)}$ 0.020 (0.013, 0.025)
	$\beta_{12MRC1}$ -7.80 (-11.20, -4.41)	-7.80 (-8.22, -7.35)	$\beta_{21MRC1}$ -0.22 (-0.47, 0.03)	-0.03 (-0.21, 0.14)	
	$\beta_{12MRC2}$ -10.78 (-11.16, -10.40)	-10.79 (-11.44, -10.13)	$\beta_{21MRC2}$ -0.37 (-0.91, 0.16)	-0.18 (-0.44, 0.06)	
	$\beta_{12SHEP}$ -8.39 (-9.15, -7.64)	-8.40 (-8.96, -7.72)	$\beta_{21SHEP}$ -0.48 (-1.01, 0.06)	-0.29 (-0.57, -0.07)	
	$\beta_{12STOP}$ -14.28 (-15.03, -13.52)	-14.28 (-15.99, -13.02)	$\beta_{21STOP}$ -0.39 (-0.94, 0.16)	-0.21 (-0.78, 0.23)	
2	$\beta_{12}$ -10.25 (-12.48, -8.01)	-10.25 (-10.84, -9.65)	$\beta_{21}$ -0.16 (-0.27, -0.04)	-0.12 (-0.25, -0.01)	$\alpha^{(2)}$ 0.019 (0.013, 0.026) $\alpha^{(3)}$ -0.036 (-0.105, 0.027)
3	$\beta_{12}$ -10.35 (-12.55, -8.16)	-2.54 (-2.88, -2.16)	$\beta_{21}$ -0.16 (-0.28, -0.04)	-0.13 (-0.25, -0.01)	$\alpha^{(2)}$ 0.019 (0.013, 0.025) $\alpha^{(3)}$ 0.034 (0.028, 0.038)
4	$\beta_{12COOP}$ -10.18 (-11.88, -8.48)	-10.20 (-12.25, -7.94)	$\beta_{21}$ -0.16 (-0.27, -0.04)	-0.12 (-0.26, -0.01)	$\alpha^{(2)}$ 0.019 (0.013, 0.025)
	$\beta_{12MRC1}$ -7.80 (-11.20, -4.41)	-7.81 (-8.23, -7.33)			
	$\beta_{12MRC2}$ -10.78 (-11.16, -10.40)	-10.78 (-11.46, -10.22)			
	$\beta_{12SHEP}$ -8.39 (-9.15, -7.64)	-8.39 (-9.00, -7.79)			
	$\beta_{12STOP}$ -14.28 (-15.03, -13.52)	-14.28 (-15.67, -13.08)			
5	$\beta_{12}$ -10.25 (-12.48, -8.01)	-10.25 (-10.76, -9.65)	$\beta_{21}$ -0.16 (-0.27, -0.04)	-0.12 (-0.24, -0.01)	$\alpha^{(2)}$ 0.019 (0.013, 0.025) $\alpha^{(3)}$ -0.036 (-0.098, 0.034)

Table 16: One-stage joint and separate model results for analysis of SBP and time to MI by model group (dataset contains 28977 individuals, 1124 events, and 157923 longitudinal measurements)

# Longitudinal Treatment Effect

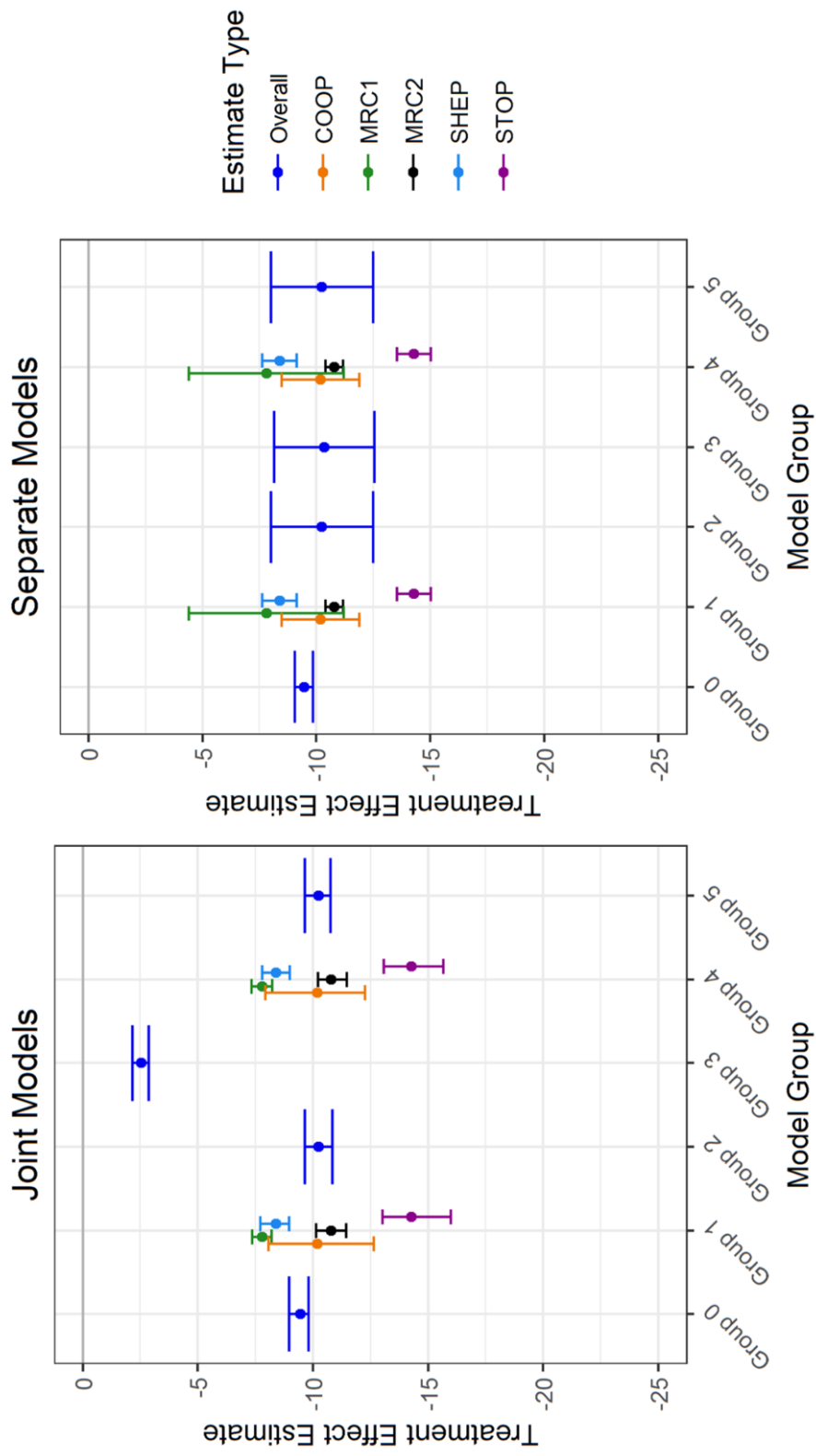


Figure 16: Graphical representation of estimates of longitudinal treatment effect shown in Table 16 for one-stage analysis of SBP and time to MI

# Time-to-Event Treatment Effect

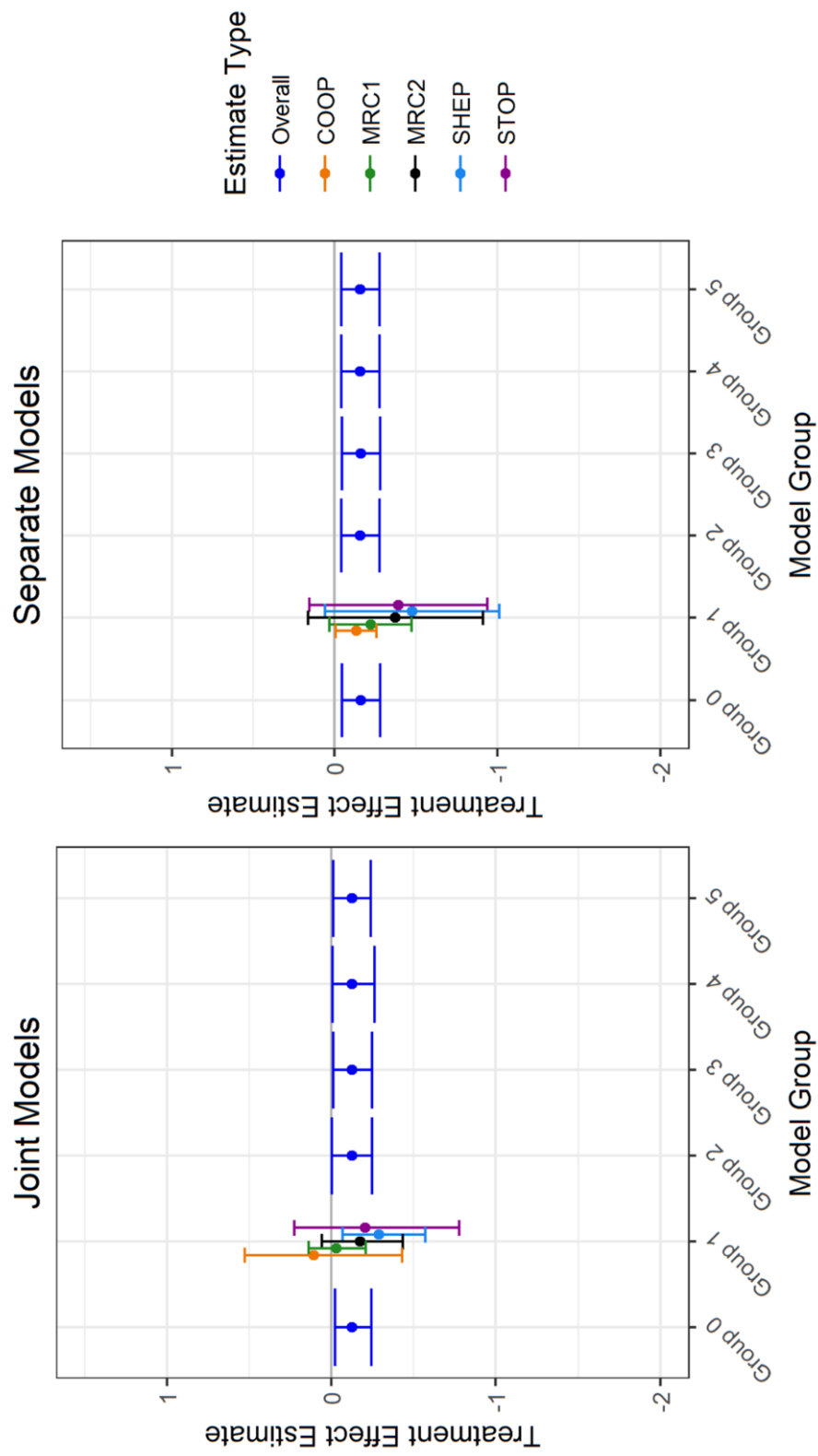


Figure 17: Graphical representation of estimates of time-to-event treatment effect shown in Table 16 for one-stage analysis of SBP and time to MI

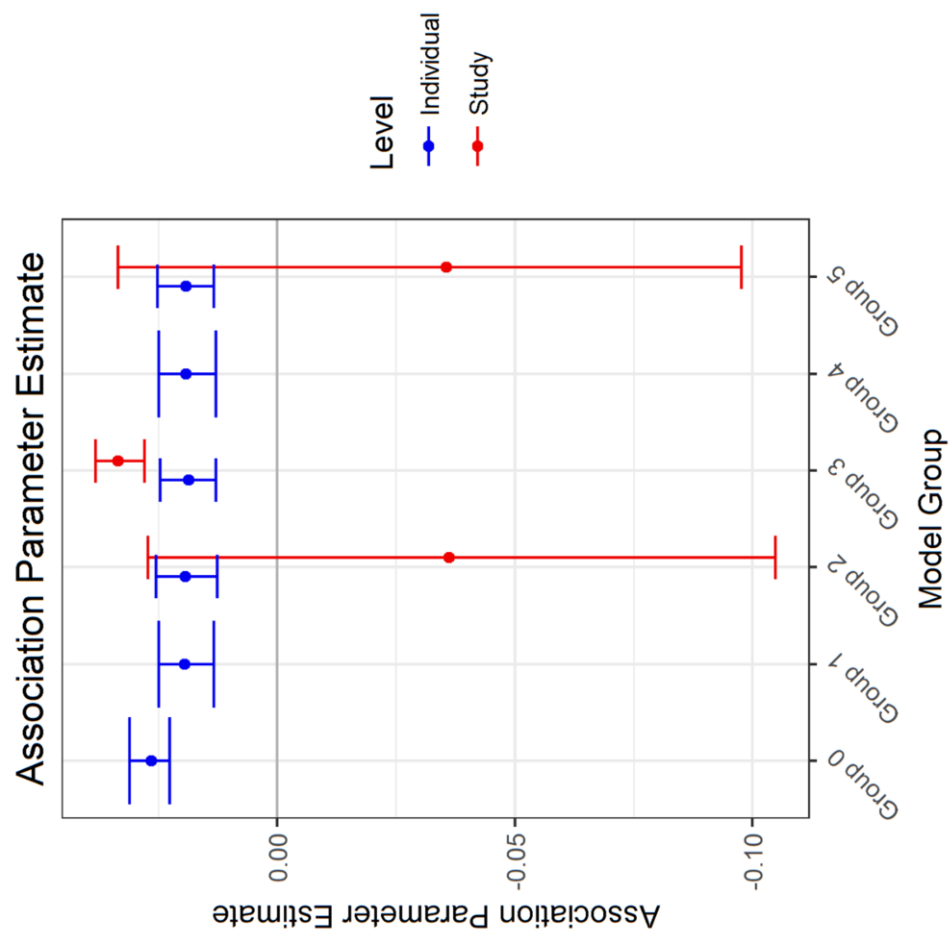


Figure 18: Graphical representation of estimates of association parameters shown in Table 16 for one-stage analysis of SBP and time to MI

### 6.3.2.3 *SBP and time to stroke*

All estimates of the longitudinal treatment effect for SBP and time to stroke across all model groups for both the joint and separate analyses were significant and negative (Table 17). As such there was evidence that assignment to any drug intervention for hypertension versus no treatment, placebo or usual care significantly reduced SBP. Effect estimates between the joint and separate methods and across model groups were similar, apart from the joint model results from group 3 (which accounted for between study heterogeneity solely using study level random effects). The joint model group 3 longitudinal treatment effect was significant and negative but smaller in magnitude than the other results. The group 3 results for the separate model were comparable to the other model groups. Again, there was variation in the size of confidence intervals between the joint and separate models (Figure 19).

The time-to-event treatment effect coefficient for SBP and time to stroke showed evidence of a significant reduction in risk of stroke for those assigned to any drug intervention for hypertension versus no treatment, placebo or usual care for the joint and separate analyses from model groups 0, 2, 3, 4, 5 (Table 17). In addition, for model group 1, a significant negative treatment effect was estimated in studies COOP, MRC1, SHEP and STOP from the joint model, and for studies COOP and MRC1 from the separate model. Effect estimates were similar across model groups, and between joint and separate models. From Figure 20, it can be seen that confidence intervals for model groups 0 and 2 through 5 are comparable between the joint and separate methods, although there are noticeable differences between the intervals for model group 1.

As with the other analyses, across all model groups, all the individual level association parameters were statistically significant and positive, although small in magnitude (Table 17, Figure 21). This indicated evidence that individuals with greater than average SBP were at higher risk of a stroke.

Again, for the study level association parameters calculated for groups 2, 3, and 5, the groups containing just a study level random intercept (2 and 5) returned an insignificant negative study level association parameter (Table 17, Figure 21). However, group 3 (which contained a study level random intercept and treatment effect) returned a significant positive study level association parameter, interpreted that studies with a study population average SBP greater than the overall population average had a higher risk of stroke.

### SBP and time to stroke

Model Group	Longitudinal Treatment Effect Parameter(s)		Time-to-Event Treatment Effect Parameter(s)		Association parameters
	Separate Model Results	Joint Sub-Model Results	Separate Model Results	Joint Sub-Model Results	
0	$\beta_{12}$	-9.43 (-9.82, -9.05)	-9.44 (-9.87, -9.08)	$\beta_{21}$	$\alpha^{(2)}$ 0.044 (0.040, 0.048)
1	$\beta_{12COOP}$	-9.98 (-11.68, -8.28)	-9.98 (-11.84, -7.76)	$\beta_{21COOP}$	$\alpha^{(2)}$ 0.034 (0.027, 0.041)
	$\beta_{12MRC1}$	-7.79 (-11.20, -4.39)	-7.79 (-8.18, -7.39)	$\beta_{21MRC1}$	
	$\beta_{12MRC2}$	-10.74 (-11.12, -10.36)	-10.74 (-11.50, -10.10)	$\beta_{21MRC2}$	
	$\beta_{12SHEP}$	-8.39 (-9.14, -7.63)	-8.40 (-9.00, -7.83)	$\beta_{21SHEP}$	
	$\beta_{12STOP}$	-14.24 (-15.00, -13.49)	-14.25 (-15.71, -12.88)	$\beta_{21STOP}$	
2	$\beta_{12}$	-10.19 (-12.41, -7.96)	-10.19 (-10.75, -9.60)	$\beta_{21}$	$\alpha^{(2)}$ 0.034 (0.026, 0.042) $\alpha^{(3)}$ -0.077 (-0.189, 0.003)
3	$\beta_{12}$	-10.29 (-12.48, -8.11)	-2.52 (-2.93, -2.15)	$\beta_{21}$	$\alpha^{(2)}$ 0.030 (0.023, 0.038) $\alpha^{(3)}$ 0.056 (0.051, 0.060)
4	$\beta_{12COOP}$	-9.98 (-11.68, -8.28)	-9.97 (-12.30, -7.48)	$\beta_{21}$	$\alpha^{(2)}$ 0.034 (0.026, 0.042)
	$\beta_{12MRC1}$	-7.79 (-11.20, -4.39)	-7.79 (-8.18, -7.36)		
	$\beta_{12MRC2}$	-10.74 (-11.12, -10.36)	-10.75 (-11.41, -10.04)		
	$\beta_{12SHEP}$	-8.39 (-9.14, -7.63)	-8.40 (-9.06, -7.69)		
	$\beta_{12STOP}$	-14.24 (-15.00, -13.49)	-14.24 (-15.65, -12.92)		
5	$\beta_{12}$	-10.19 (-12.41, -7.96)	-10.19 (-10.70, -9.66)	$\beta_{21}$	$\alpha^{(2)}$ 0.034 (0.027, 0.041) $\alpha^{(3)}$ -0.076 (-0.171, 0.005)

Table 17: One-stage joint and separate model results for analysis of SBP and time to stroke by model group (dataset contains 28985 individuals, 808 events, and 157834 longitudinal measurements)

# Longitudinal Treatment Effect

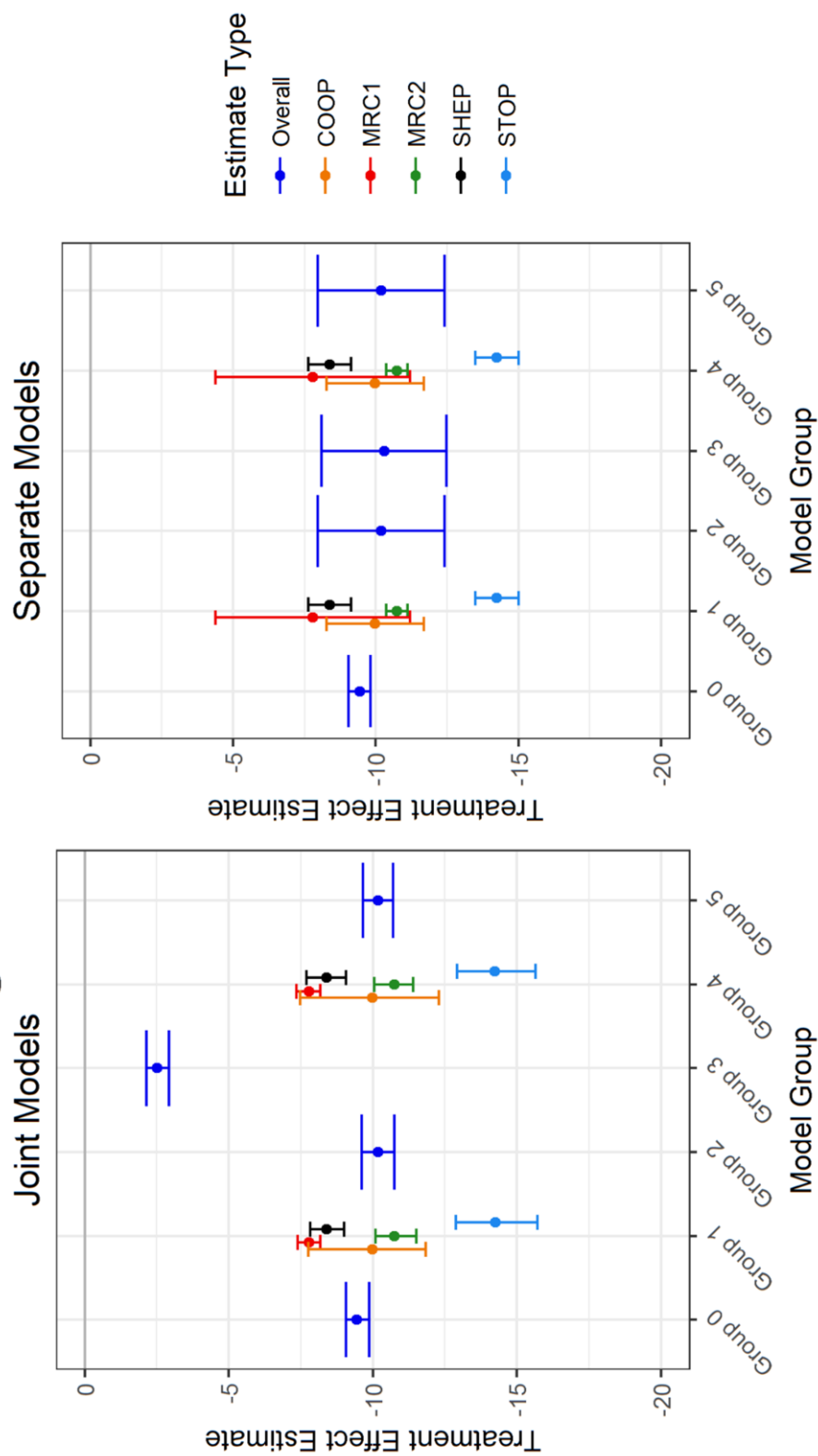


Figure 19: Graphical representation of estimates of longitudinal treatment effect shown in Table 17 for one-stage analysis of SBP and time to stroke

# Time-to-Event Treatment Effect

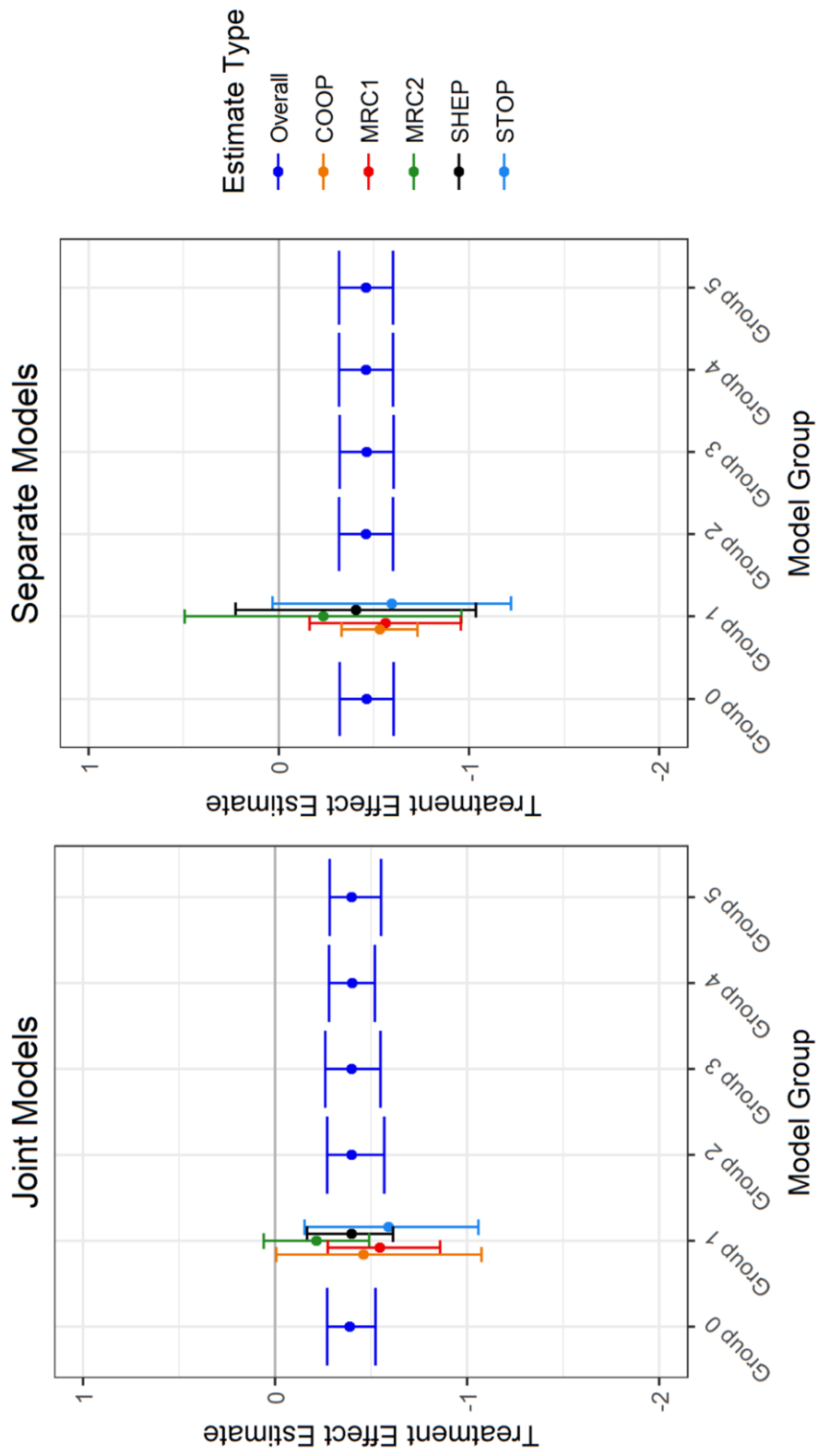


Figure 20: Graphical representation of estimates of time-to-event treatment effect shown in Table 17 for one-stage analysis of SBP and time to stroke



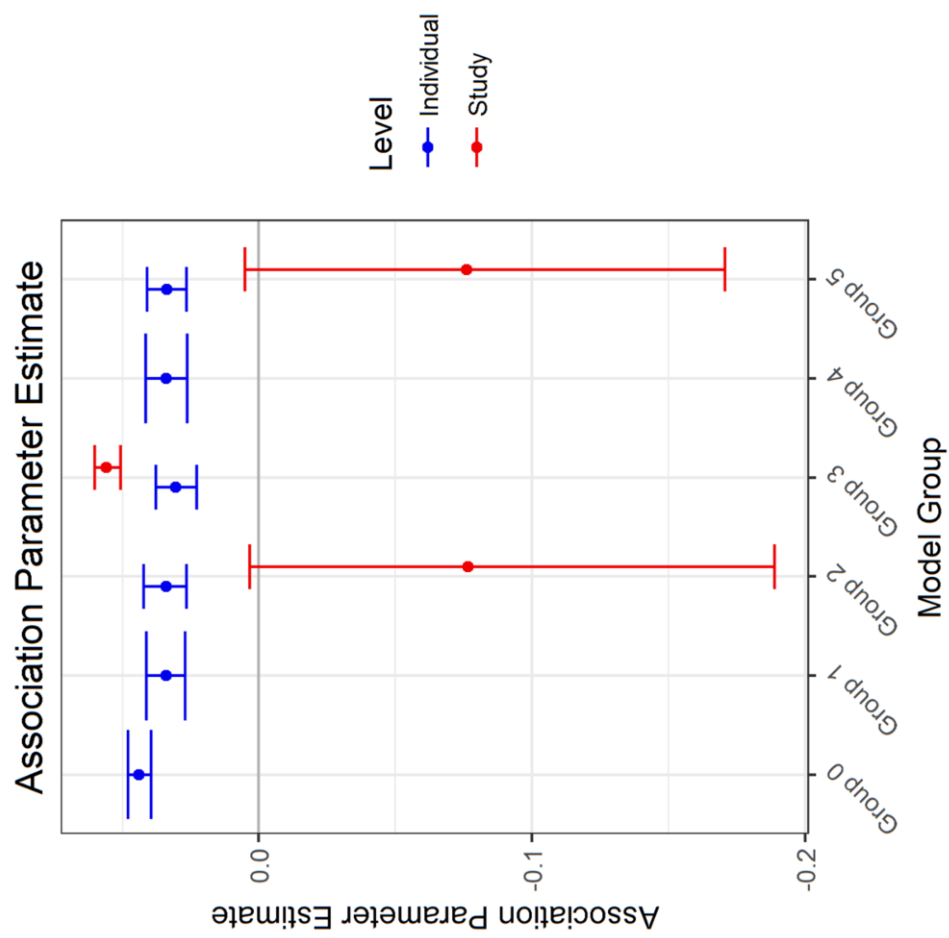


Figure 21: Graphical representation of estimates of association parameters shown in Table 17 for one-stage analysis of SBP and time to stroke

## 6.4 Discussion of Joint Meta-Analysis of Real Data

During this chapter a demonstration of the methods discussed in this thesis applied to the INDANA has been presented. Both a two-stage and a one-stage meta-analysis of the dataset have been conducted.

### 6.4.1 Discussion of Two-Stage joint MA of INDANA data

In the two-stage analysis of the INDANA dataset, based on the joint modelling analysis, assignment to any treatment for hypertension versus no treatment, placebo or usual care significantly reduced SBP across all studies examined (heterogeneity across studies was present in the magnitude of the effect, but the direction and significance of estimates is consistent across studies). Assignment to any treatment for hypertension, versus no treatment, placebo or usual care, had no significant effect on risk of an event (based on random effects MA given the heterogeneity of estimates, although assignment to any treatment for hypertension appears to have a greater (but still insignificant) effect on risk of stroke, than on MI or death).

Evidence of heterogeneity was noted between studies for several of the parameters, such as the association parameter (where the estimate from MRC1 often differed from the remaining studies), and longitudinal treatment effect estimate. This could be attributable to the fact that the studies differed in demographics known to be linked to the disease area (Section 6.1.1). The Cochrane handbook [184] (section 9.5.3) details seven approaches to dealing with heterogeneous parameter estimates in the second stage of two stage MA (or AD-MA), namely checking the data is correct, not pooling results in a MA, exploring heterogeneity (through a subgroup analysis or meta-regression), ignoring the heterogeneity (as the p value from a fixed effect MA is a valid test of the null hypothesis of no effect in every study, although ideally heterogeneity should be explained and accounted for), performing random effects MA in place of fixed effects MA, changing the effect measure (such as using standardized results), or finally by excluding studies. Alternatively, given the availability of IPD, two or one-stage analyses could be conducted, including of demographic characteristics thought to influence the effect measure in the sub-models, or even employing more complex association structures such as ones that allow interaction between association and certain covariates.

Various tests and procedures exists for assessing the level of heterogeneity when pooling parameters including the p value for the  $\chi^2$  test for presence of heterogeneity in treatment

effects (which is low powered where included studies are few or small in size), the estimate of between study heterogeneity  $\tau^2$  (which takes larger values for analyses with greater heterogeneity, but is not a test statistic that can be compared to certain criteria) or the  $I^2$  statistic, which states the percentage of variability in effect estimates due to heterogeneity rather than chance [184]. The values for  $I^2$  have been grouped into four overlapping categories, namely values of 0% to 40% represent potentially unimportant heterogeneity, 30% to 60% moderate heterogeneity, 50% to 90% substantial heterogeneity, and 75% to 100% considerable heterogeneity. Overall, presence of heterogeneity is often a diagnosis made by the researcher, based on the combination of evidence from the p value,  $\tau^2$ ,  $I^2$ , and visual assessment of the forest plot (e.g. non-overlap of confidence intervals between studies could indicate heterogeneity), rather than based on the result of a statistical test. As this investigation aimed to demonstrate methods, rather than conduct an in-depth assessment of potential treatment modifiers, covariates such as age, gender etc. have not been included in the examined models.

In this real example of two-stage meta-analysis of joint data, there was mostly little difference between the estimates obtained using separate longitudinal or time-to-event models in the first stage compared to using joint models (apart from the difference in statistical significance between the pooled estimates for the time-to-event treatment effect coefficient between separate and joint models). This agreement between separate and joint models is expected, as the association parameters from the random effects MA (examined due to presence of heterogeneity) were not statistically significant, and all association parameter estimates were small in magnitude.

In single study cases, evidence has been presented indicating that less biased results are obtained by using joint models rather than separate models in cases where the longitudinal and the time-to-event outcomes are correlated [158]. To investigate further whether this behaviour persists in the multi-study case, a two-stage MA simulation study was performed examining scenarios with a range of magnitudes of association parameter (presented in Chapter 7).

#### 6.4.2 Discussion of One-Stage joint MA of INDANA data

Throughout the one-stage real data investigation, all analyses from all model groups for both joint and separate models estimated significant negative longitudinal treatment effect. This is interpreted that assignment to any drug intervention for hypertension versus no treatment, placebo or usual care significantly reduces SBP. It was expected that the

longitudinal estimates across all pairwise combinations of outcomes (SBP and time to death, SBP and time to MI, and SBP and time to stroke) are similar, as the longitudinal outcome is the same.

The estimates of the longitudinal treatment effect produced by the joint models fitted under the group 3 specification (see Table 15-Table 17) differed significantly from those produced by the other groups. This may be attributable to the fact that the group 3 models accounted for between study heterogeneity solely through study level random effects, whose distribution is based effectively on a number of data points equal to the number of studies involved in the analysis. As this was small (5 or 6) in all the analyses conducted, the distribution may be badly estimated. The other groups that employed study level random effects (groups 2 and 5) only included one study level random effect, and employed other additional methods such as fixed terms or stratified baseline hazard to account for between study heterogeneity. The estimates of longitudinal treatment effect from these groups were more similar to the groups that involved no study level random effects. This may be due to a variety of reasons e.g. a simpler study level random effects distribution to estimate. It is important to determine whether the estimates provided by joint models of a group 3 type specification become more reliable as the number of studies contributing to the meta-analysis increases. This hypothesis is investigated in the one-stage simulation study described in Section 7.2.

It is interesting to note that the longitudinal treatment effect estimate from group 3 for the separate model was similar to the results from the remaining groups for both the separate and joint models. This may be attributable to the fact that in the separate model, the study level random effects in group 3 solely model the between study variation in longitudinal intercept and treatment effect. However, in the joint model, they are additionally involved in modelling the link between the longitudinal and time-to-event outcomes, and accounting for between study heterogeneity in the time-to-event sub-model. The simulations in Section 7.2 will investigate whether this disparity in group 3 models between separate and joint analyses persists under a range of conditions.

Throughout, the results from separate and joint models have been similar, although in several cases for the time-to-event treatment effect estimate, the joint model estimate has been closer to zero. The similarity between the methods may be due to the fact that whilst the association parameters are generally significant, they are small in magnitude. Differences between separate and joint models may be observed with significant

association parameters, a scenario investigated in the one-stage simulation study (Section 7.2). However, effect of the link between the longitudinal and time-to-event outcomes might be better assessed through the magnitude of the  $W_{2ki}(t)$  term (equal to  $\alpha^{(2)}(Z_{ki}^{(2)}b_{ki}^{(2)}) + \alpha^{(3)}(Z_{ki}^{(3)}b_k^{(3)})$ , with  $\alpha^{(3)}(Z_{ki}^{(3)}b_k^{(3)})$  removed if study level random effects are not included in the model). As such, the value of  $W_{2ki}(t)$  was output for each individual, for each study in the one-stage analysis as well as overall across all studies, for each of SBP and time to death, SBP and time to MI, and SBP and time to stroke. These results are available in Appendix 5. The median, lower and upper quartiles, and minimum and maximum observed values have been reported in Table 29, and the densities of the  $W_{2ki}(t)$  terms summarised graphically in Figures 60-62. From these results, several points can be noted. Firstly, there is often a noticeable difference between results from model group 0 (which ignores between study heterogeneity) and the remaining model groups, highlighting the importance of accounting for between study heterogeneity when it is known or suspected. Secondly, the results from model group 3 are often different to those from model groups 1, 2, 4 or 5 (as observed with the parameter estimates already reported in this chapter). However the behaviour of the  $W_{2ki}(t)$  produced by model group 3 differs between studies. For studies COOP and STOP the model group 3  $W_{2ki}(t)$  are higher in magnitude than those from model groups 1, 2, 4 or 5, whereas for MRC1 and SHEP they are lower. For MRC2, and (for SBP and time to death) EWPHE, the  $W_{2ki}(t)$  estimates are similar between model groups 3 and model groups 1, 2, 4 and 5. This variability might suggest that a model that allows for heterogeneity in the association parameter might be beneficial (this is mentioned as planned future extension in Chapter 8).

The estimates of  $W_{2ki}(t)$  across all three sets of analyses (time to death, MI or stroke) show densities that include 0, although the estimates from SBP and time to stroke are not as closely centred about zero as for SBP and time to death, or SBP and time to MI. As such, this examination of  $W_{2ki}(t)$  reinforces the conclusion that these analyses do not show a strong, consistent relationship between SBP and the events of interest.

Figures 15, 18 and 21, which display the association parameter estimates and 95% CIs for the analyses of SBP and each of time to death, time to MI and time to stroke, display a pattern in the estimate of the study level association parameter  $\alpha^{(3)}$  for model groups 2, 5 and 3. The estimates produced by model groups 2 and 5 are in agreement; close to zero, non-significant, with wide confidence intervals. However the result produced by model group 3 is significant and positive, with narrow confidence intervals. This estimate of  $\alpha^{(3)}$

produced by model group 3 should be treated with caution, as it is inconsistent with the results produced by model groups 2 and 5 (this behaviour holds for other parameter estimates; for example, model groups 2 and 5 produce longitudinal treatment effect estimates similar to those produced by the other model groups, whereas model group 3 produces estimates inconsistent with those produced by the other model groups). As such, there is some suggestion that model group 3 does not reliably estimate model parameter estimates from the sub-models or association structure. As mentioned before, this may be due to the study level random effects being poorly estimated.

Issues with estimation of study level random effects have been stated as a potential reason for the differing performance of model group 3. The study level random effects quantify the difference between study populations in the dataset. A large number of level 2 units (total individuals included in the meta-analysis) aids with the estimation of the distribution of individual level random effects. A large study population helps to establish the behaviour and demographic of that single study. However, the level of variability between studies is assessed by comparison between different study populations, and so is influenced by the number of studies included, rather than the sample size contained within each study. The narrow confidence intervals displayed for e.g. the  $\alpha^{(3)}$  parameter from model group three could be attributable to the large number of level 2 units (individuals in the meta-analysis), but the small number of level 3 units (studies in the meta-analysis) could lead to the point estimates themselves being incorrect (potentially explaining the difference in parameter estimates reported by model groups 3 compared to the other model groups).

In this demonstration of the methodology developed in Chapter 3, we have fitted examples from a range of approaches to account for between study heterogeneity. Fitting a range of models has allowed us to highlight the inconsistency between model group 3 (which solely accounts for between study heterogeneity using study level random effects) and the remaining model groups. In practice, a one-stage meta-analysis may not employ several approaches to account for between study heterogeneity. However, it might be recommended that, in complex MA such as those for joint data, in a secondary analysis, the primary analysis be repeated using an alternative method to account for between study heterogeneity, as if the results produced by the secondary analysis are consistent with the main analysis, confidence in the reliability of the main analysis would be increased.

#### 6.4.3 Comparison of one and two-stage approaches

There were similarities between the results of two and the one-stage analyses. The longitudinal treatment effect was consistently statistically significant and negative across all analyses regardless of the methods used. As such, there is clear evidence that allocation to any hypertensive medication versus no treatment, placebo or usual care significantly reduces SBP.

The time-to-event treatment effect estimates were also similar between the one and two-stage analyses, however there were differences in significance of results. The two-stage analysis showed a significant reduction in risk of an event for SBP and time to stroke based on the joint analysis, whereas the one-stage analysis showed an additional significant reductions in risk of an event for SBP and time to MI. Furthermore, in the two-stage analysis, the individual level association parameters showed evidence of heterogeneity, with pooled non-significant estimate from the random MA. Conversely, the one-stage analysis produced significant positive (small magnitude) individual level association parameters across all analyses conducted. Overall, whilst the point estimates obtained from the two-stage and one-stage analyses were similar, there were differences in significance of results.

Recently Burke et al [189] described the common reasons that results from one and two-stage analyses based on the same data may differ. In our case, the discrepancy might be attributable to the fact that one-stage methods used exact likelihood methods versus, the two-stage which used approximate likelihood methods (which Burke et al note could be unreliable for time-to-event outcomes when the outcome is rare). Another potential reason for the discrepancy may be that whilst the one-stage approach accounts for correlation between parameters, standard MA methods in a two-stage MA do not. A solution to this could be to utilise multivariate meta-analytic methods in the second stage of a two stage MA, discussed in Section 8.3.2.1.

#### 6.4.4 Limitations of conducted analyses

These investigations focused only on demonstrating the methods contained in this thesis, and did not include some variables known to be linked to CVD. Inclusion of these variables may reduce some of the heterogeneity seen. Future analyses aiming to influence healthcare should investigate in detail the covariates known to effect hypertension or CVD.

The analyses conducted in this chapter were restricted to joint models that employed a linear mixed effects model for the longitudinal sub-model, with a PH model for the time-to-

event sub-model, with zero mean random effects with level specific association parameters forming the association structure. However, other joint modelling specifications may be preferable, such as use of splines in the longitudinal sub-model to account for the mentioned change in trajectory at around 6 months. In addition, given that the Kaplan-Meier curves cross for some outcomes for some studies investigation of other time-to-event sub-model types (such as AFT or parametric models) may prove beneficial. Finally, clinically, it may make sense to investigate alternative association structures, e.g. current value structure that would model how the true recorded value of the longitudinal outcome affects the risk of an event. The analyses were restricted to the presented model specifications by the choice of software used, however there are plans in the future to expand the *joinerMeta* package to allow for greater flexibility in model structure (see Section 8.3.2.2).

In addition, to account for the change in slope in the longitudinal trajectories, an  $\exp(-3 * t)$  term was included in both the one and two-stage analyses. However a random effect was only included for the linear time term (whose interpretation may be complicated if the effect of the  $\exp(-3 * t)$  differs across individuals). As such, the analysis may be improved by investigation of more complex random effects structures, such as assignment of an individual level random effect to  $\exp(-3 * t)$ . This analysis aimed only to demonstrate the methodology produced in this research. In an analysis aiming to influence healthcare, an in depth investigation of potential fixed and random effect specifications would be undertaken. The results in the real data analysis display some heterogeneity between studies. Here, the heterogeneity is likely to be due to the differences in demographics between studies included in the meta-analysis, as the included studies differ in parameters known to be linked to the outcomes, such as age (see Section 6.1). An adjusted analysis may display reduced heterogeneity. This is reinforced by the fact that the simulation studies displayed less heterogeneous results (see Chapter 7, although it should be noted that simulated data is generally less heterogeneous than real data, for example due to the lack of measurement error).

#### 6.4.5 Concluding remarks

Overall, this chapter has demonstrated the application of methods developed in this thesis to a real dataset. The analyses shown here motivated both the one and two-stage simulation studies described in Chapter 7, and the proposed further work to be conducted, discussed in Chapter 8.



## Chapter 7 : Simulation investigation of joint meta-analytic methods

---

In this Chapter the simulation investigations undertaken both for the two-stage and the one-stage MA of joint longitudinal and time-to-event data are described.

The INDANA dataset displayed small magnitude associations between the longitudinal and time-to-event outcomes, resulting in little difference between joint and separate MA (Chapter 6). However, joint models have shown benefits (such as reduced bias and increased efficiency) over separate analyses for single study cases [158]. As such, it is necessary to establish if joint models show benefit over separate models in MA under a range of association magnitudes. Simulations to investigate this for one and two-stage MA approaches are shown in Sections 7.1 and 7.2.

Additionally, the one-stage INDANA analyses highlighted potential issues with the proposed one-stage models that involved study level random effects. Consequently, to determine what conditions are required for estimation of study level random effects to be reliable, additional one-stage simulation studies were conducted that varied the number of studies included in the MA, and varied the level of between study heterogeneity (see Section 7.2).

During the simulation studies, the University of Liverpool's HTCondor system was used, see [240], <https://research.cs.wisc.edu/htcondor/> , and <http://condor.liv.ac.uk/> . The condor system at the university makes use of over 500 computers across the university campus, and allows "jobs" such as R analyses to be run on any idle computers. Such systems allow analyses that require the same procedure to be carried out multiple times, such as in simulation studies, to be completed sooner by running the analyses in parallel on different machines. The system has restrictions, for example analyses cannot run for over 24 hours as the system resets daily, and partially completed analyses can be forced to restart if the computer they are running on becomes unavailable. However, the system is a powerful tool to allow in depth simulations studies to take place.

### 7.1 Simulations investigating two-stage methods

The two-stage MA simulation investigation described here was reported as part of Sudell et al [176]. In the following sections, the background methodology of the data simulation is described. Analysis of the simulated data follows the method described in Section 3.2. The

models fitted to the simulated data are defined, and the results of the simulation study presented and discussed.

### 7.1.1 Simulation of joint data for two-stage simulations

During the simulation study, joint data containing a single continuous normally distributed longitudinal outcome and a single censored time-to-event outcome was simulated, using the *simjoint* data simulation function in the *joiner* package [80]. Each dataset contained 5 studies, each with 500 individuals randomised equally to two treatment groups. A maximum of 5 longitudinal measurements at times 0, 1, 2, 3, 4 were permitted, with measurements retained only up to the individual's survival time. Data for each study was simulated separately, and then pooled to form a meta-dataset. The joint data was simulated under the model shown in equation (67).

$$\begin{aligned} Y_{kij} &= \beta_{10k} + \beta_{11k}t_{kij} + \beta_{12k}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{21k}treat_{ki} + W_{2ki}(t)) \\ W_{2ki}(t) &= \alpha_k^{(2)}W_{1ki}(t) = \alpha_k^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) \end{aligned} \quad (67)$$

In equation (67), the longitudinal outcome  $Y_{kij}$  is simulated under a model that contains fixed intercept, time and treatment terms (with corresponding coefficients  $\beta_{10k}$ ,  $\beta_{11k}$  and  $\beta_{12k}$ ), individual level random intercept and time terms ( $b_{0ki}^{(2)}$  and  $b_{1ki}^{(2)}$  respectively), and an error term  $\varepsilon_{kij}$ . The individual level random effects follow a multivariate normal distribution  $\mathbf{b}_{ki}^{(2)} \sim N(0, \mathbf{D}_k)$  and are considered independent of the errors, which are considered IID and follow a  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$  distribution. The longitudinal and time-to-event outcomes were linked through a proportional random effects only association structure, where the individual level random effects were inserted with common coefficient  $\alpha_k^{(2)}$  (with event times  $T_{Ski}$  replacing longitudinal times  $t_{kij}$ ) into the time-to-event sub-model, giving term  $W_{2ki}(t)$ . The coefficients for the fixed effects in both the longitudinal and time-to-event sub-models were common across simulated studies.

The time-to-event outcome was simulated using the methodology described by Bender et al [219] and Austin [220], an overview of the implementation of their methods for this simulation study follows. Event times were considered to be Gompertz distributed, with scale parameter  $\kappa = \exp(\theta_0 + \alpha_k^{(2)}b_{0ki}^{(2)})$ ,  $\kappa > 0$ , and shape parameter  $\omega = \theta_1 + \alpha_k^{(2)}b_{1ki}^{(2)}$ ,  $-\infty < \omega < \infty$ . In the shape and scale parameter definitions,  $\alpha_k^{(2)}$  is the common association parameter for the individual level random effects,  $\theta_0$  and  $\theta_1$  are parameters used to control the distribution of the event time (which are specified by the user in the

function written to generate the event times), and  $b_{0ki}^{(2)}$  and  $b_{1ki}^{(2)}$  are respectively the individual specific random intercept and time terms. The extreme value distribution was used to calculate the values to assign to  $\theta_0$  and  $\theta_1$ . Using that the random effects  $b_{ki}^{(2)}$  had expectation of zero, for a Gompertz distribution:

$$\mathbb{E}(T) = \mu_0 = -\frac{1}{\omega} \left( \log\left(\frac{\kappa}{\omega}\right) + \gamma \right)$$

$$Var(T) = \sigma_0^2 = \frac{\pi^2}{6\omega^2}$$

In the above equations,  $\pi \approx 3.142$ , the mathematical constant pi, and  $\gamma \approx 0.5772$  is Euler's constant. A mean event time of  $\mu_0 = 3$  with standard deviation  $\sigma_0 = 1$  was specified. Given that longitudinal time points were simulated as 0, 1, 2, 3, 4, through testing 10,000 realisations from a Gompertz distribution with these specifications, it was confirmed that this distribution would produce individuals with varying numbers of recorded longitudinal measurements. The above equations were rearranged in terms of  $\omega$  and  $\kappa$ , to give:

$$\omega = \frac{\pi}{\sqrt{6}\sigma_0} = \frac{\pi}{\sqrt{6}}$$

$$\kappa = \omega \exp(-\gamma - \mu_0\omega) = \frac{\pi}{\sqrt{6}} \exp\left(-\gamma - \frac{3\pi}{\sqrt{6}}\right)$$

As such, given the definitions of the shape and scale parameter:

$$\theta_0 = \log(\kappa) = \log\left(\frac{\pi}{\sqrt{6}} \exp\left(-\gamma - \frac{3\pi}{\sqrt{6}}\right)\right)$$

$$\theta_1 = \omega = \frac{\pi}{\sqrt{6}}$$

These values were then supplied as  $\theta_0$  and  $\theta_1$  to the data simulation code.

The event rate (proportion of individuals in the data in each study that experienced an event) was controlled through the censoring times (which followed an exponential distribution, with parameter  $\varphi$ ). Two sets of data were produced, one with a “high” event rate (~75%) and one with a “low” event rate (~25%). A range of  $\varphi$  parameters for the censoring distribution were tested given the distribution of the event times, resulting in  $\varphi = \exp(-3.08)$  and  $\varphi = \exp(-0.58)$  for the high and low event rate data respectively.

As the longitudinal data being simulated contained both a random intercept and slope, to ensure presence of censorings, individual's event times were generated in a two-stage

process. Firstly a random number  $U_{ki}$  was generated from a uniform  $U(0,1)$  distribution for each individual  $i$  in the dataset for study  $k$ . An indicator variable took value true if both the following conditions were satisfied, false otherwise.

$$\begin{aligned} \text{Condition 1: } & (\theta_1 + \alpha_k^{(2)} b_{1ki}^{(2)}) < 0 \\ \text{Condition 2: } & U_{ki} < \exp\left(\frac{\exp(\theta_0 + \alpha_k^{(2)} b_{0ki}^{(2)})}{\theta_1 + \alpha_k^{(2)} b_{1ki}^{(2)}}\right) \end{aligned}$$

If the indicator was set to true, the individual's event time was set to infinite (guaranteeing that they were censored). If false, a finite event time was calculated using equation (68).

$$T_{Eki} = \frac{1}{(\theta_1 + \alpha_k^{(2)} b_{1ki}^{(2)})} \log \left[ 1 - \frac{(\theta_1 + \alpha_k^{(2)} b_{1ki}^{(2)}) \log(U_{ki})}{\exp(\theta_0 + \alpha_k^{(2)} b_{0ki}^{(2)}) \exp(\beta_{21k} x_{21ki})} \right] \quad (68)$$

In equation (68),  $T_{Eki}$  represents the event time of the  $i$ th individual in the  $k$ th study,  $x_{21ki}$  represents their treatment assignment covariate, and  $\beta_{21k}$  the fixed treatment assignment coefficient in the time-to-event sub-model. The other parameters have been defined earlier. If time-stationary parameters additional to treatment assignment were included in the model, these would be included in a linear combination with  $\beta_{21k} x_{21ki}$  e.g.

$$\beta_{21k} x_{21ki} + \beta_{22k} x_{22ki} + \dots$$

As mentioned, each individual's censoring time was generated under an exponential distribution with parameter  $\varphi$  (see Bender et al [219]):

$$T_{Cki} = -\frac{\log(U_{ki})}{\varphi} \quad (69)$$

In equation (69),  $T_{Cki}$  is the censoring time for individual  $i$  in study  $k$ . The final event time for each individual is then  $T_{Ski} = \min(T_{Eki}, T_{Cki})$ , with censoring variable set to 1 if  $T_{Ski} = T_{Eki}$ , 0 otherwise.

The parameters for the models the data were simulated under were chosen to represent large differences between treatment groups, in order that deviations of different methods from the true parameter values could be clearly identified. In the longitudinal sub-model for each study  $k$ , the population intercept  $\beta_{10k}$  was set to 1, the time coefficient  $\beta_{11k}$  set to 3 and the treatment assignment coefficient  $\beta_{12k}$  was set to 2. The individual level random effects were generated under the following distribution (with the same covariance matrix for each study such that  $\mathbf{D}_k = \mathbf{D}$ ):

$$\mathbf{b}_{ki}^{(2)} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.9 & 0.5 \\ 0.5 & 1.2 \end{pmatrix}\right)$$

The IID measurement errors  $\varepsilon_{kij}$  were generated under a  $N(0, 0.01)$  distribution. The time-to-event treatment assignment coefficient  $\beta_{21k}$  for each study  $k$  was set to 3.

In order to compare joint models to separate longitudinal or time-to-event models in a range of settings, 5 association levels were investigated by setting the association parameter  $\alpha_k^{(2)}$  for each study  $k$  to 0, 0.25, 0.5, 0.75 and 1. Only positive associations were examined, as behaviour for negative associations was expected to be similar, but opposite, in direction.

For each scenario, datasets with homogenous treatment effects between studies and datasets with heterogeneous treatment effects between studies were generated. This was achieved by setting the longitudinal treatment effect coefficient  $\beta_{12k}$  to 2 and the time-to-event treatment effect coefficient  $\beta_{21k}$  to 3 for the homogenous scenarios for each study  $k$ , and instead using study specific realisations from an  $N(2, 0.5)$  and a  $N(3, 0.5)$  distribution for the heterogeneous scenarios. There was greater variation in event rates between simulated data for different levels of association in the high event rate group compared to the low event rate group (observed event rates available in Table 28, Appendix 3).

For each scenario (combination of association level, homogenous or heterogeneous treatment effect between studies, low or high event rate) 1000 datasets were simulated.

### 7.1.2 Two-stage models fitted to simulated joint data

Following methods introduced in Chapter 3, a joint model of specification shown in equation (70) was fitted to each study's data within each dataset within each scenario:

$$\begin{aligned} Y_{kij} &= \beta_{10k} + \beta_{11k}t_{kij} + \beta_{12k}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{21k}treat_{ki} + W_{2ki}(t)) \\ W_{2ki}(t) &= \alpha_k^{(2)}W_{1ki}(t) = \alpha_k^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) \end{aligned} \tag{70}$$

This model employed a linear mixed effects model for the longitudinal sub-model containing a fixed intercept, time and treatment assignment term, and individual specific random intercept and time term as well as an error term. The individual random effects in each study are assumed to follow a multivariate normal distribution with zero mean, and are considered independent of the error terms. The error terms are considered to be IID across time points and follow a zero mean normal distribution. The time-to-event sub-

model uses an unspecified baseline hazard, and contains only a fixed treatment assignment term. The association structure linking the sub-models is a proportional random effects only structure, which shares the zero mean individual level random effects between the sub-models with common association parameter  $\alpha_k^{(2)}$ . Instances of longitudinal time  $t_{kij}$  in the association structure are replaced with the survival time  $T_{ski}$ . As well as joint models, separate longitudinal and time-to-event models were fitted to each dataset. The longitudinal models had the same specification as the longitudinal sub-model of the joint model. The time-to-event models had the same specification as the time-to-event sub-model, except the  $W_{2ki}(t)$  term was not present.

Once the models were fitted, study specific model parameter estimates for the treatment effects from each sub-model ( $\beta_{12k}, \beta_{21k}$ ) and the association parameter ( $\alpha_k^{(2)}$ ) were extracted along with estimates of their standard errors, and the sample size of each study. Separate meta-analyses were performed for each parameter of interest (using the meta-analytic techniques discussed in Section 3.2.3), to produce pooled estimates  $\hat{\beta}_{12}$ ,  $\hat{\beta}_{21}$ , and  $\hat{\alpha}^{(2)}$ . Both fixed and random effects meta-analyses were conducted, and the pooled point estimates and 95% confidence intervals for the parameters extracted from each, along with heterogeneity measure  $\tau^2$ . The mean pooled estimates for each parameter, along with their standard error (the standard deviation of the pooled estimates produced by the simulation study) were recorded. During the simulation study the number of failed fits (where the model failed for fit for at least one study in the simulation run) per scenario were recorded. Coverage was also calculated for each set of simulations, as the percentage of simulation runs where the confidence interval for the pooled parameter estimates contained the “true” value the parameter was simulated under.

### 7.1.3 Two-stage simulation study results

Results for the simulation study for two-stage MA of joint data are presented in Table 18-Table 21. Plots for these results are given in Figure 22-Figure 24, with the dotted line in the mean pooled estimate column identifying the “true” value of the parameter, and in the coverage column identifying 95% coverage.

Throughout the scenarios investigated, the fixed and random effect MA results were similar for the homogenous datasets, while the random effect MA showed better coverage for the heterogeneous datasets. Consequently, the plots show only the results from the random MA, with results from both the fixed and random MA available in Tables 2-5. Additionally,

the estimates of  $\tau^2$  were larger (as expected) for the heterogeneous than the homogenous data.

Across scenarios, regardless of the level of association between the longitudinal and the time-to-event outcomes, the mean pooled estimates of the longitudinal treatment effect coefficient ( $\hat{\beta}_{12}$ ) based on joint model fits and those based on the separate longitudinal models agreed well (Table 18-Table 21, Figure 22).

The mean pooled estimates for the treatment effect coefficient for the time-to-event outcome ( $\hat{\beta}_{21}$ ) based on use of joint models in the first stage of the two-stage MA are close to the true values of the coefficients across all tested association levels (Table 18-Table 21, Figure 23).

Conversely, different behaviour is observed for the mean pooled estimates of  $\hat{\beta}_{21}$  from analyses that use a separate time-to-event model in the first stage of the MA (Table 18-Table 21, Figure 23). The results agree with those based on the joint model when association is 0, across all scenarios tested. However, once association is non-zero, the mean pooled estimate from the separate model increasingly underestimates the true parameter value as the magnitude of individual level association increases, with coverage dropping to zero. Also the empirical standard errors of the pooled  $\hat{\beta}_{21}$  estimates from the joint models appear relatively constant, with some increase with increasing association. However, as association increases in magnitude, the empirical standard errors of the pooled  $\hat{\beta}_{21}$  estimates from the separate time-to-event approach decreases.

The association parameter is only produced by the joint model approach. Across all simulation scenarios, the mean pooled  $\hat{\alpha}^{(2)}$  estimates were close to the true value (to be expected given that the joint models utilise the same association structure that the data was generated under). The empirical standard error of the pooled estimate increased slightly as association increases. The coverage also decreased slightly for larger magnitudes of association (Table 18-Table 21, Figure 24).

Two-Stage Simulation Results						
Homogenous treatment effect between studies, low event rate data						
Model	Scenario					
	$\alpha^{(2)} = 0$	$\alpha^{(2)} = 0.25$	$\alpha^{(2)} = 0.5$	$\alpha^{(2)} = 0.75$	$\alpha^{(2)} = 1$	
Joint longitudinal and time-to-event model (JointR, proportional association, shared random effects, unspecified baseline hazard)						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.00 (0.04) [94.9]	2.00 (0.04) [94.3]	2.01 (0.04) [92.6]	2.02 (0.04) [92.9]	2.02 (0.04) [89.4]
	Random MA $\tau^2$	2.00 (0.04) [96.5]	2.00 (0.04) [95.7]	2.01 (0.04) [94.4]	2.02 (0.04) [94.1]	2.02 (0.04) [91.8]
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.92 (0.15) [92.1]	2.94 (0.14) [93.6]	2.96 (0.13) [94.1]	2.95 (0.13) [92.7]	2.93 (0.14) [89.2]
	Random MA $\tau^2$	2.93 (0.14) [94.2]	2.95 (0.14) [95.2]	2.97 (0.13) [95.7]	2.96 (0.13) [94.8]	2.93 (0.14) [91.3]
Association estimate (true $\alpha^{(2)}$ as in column headers)	Fixed MA	0.05 (0.11)	0.08 (0.12)	0.09 (0.13)	0.09 (0.13)	0.09 (0.13)
	Random MA $\tau^2$	0.00 (0.01) [93.8]	0.25 (0.02) [95.6]	0.50 (0.02) [93.2]	0.74 (0.03) [93.1]	0.98 (0.03) [88.9]
	Random MA	0.00 (0.01) [95.9]	0.25 (0.02) [97.1]	0.50 (0.02) [95.9]	0.74 (0.03) [94.7]	0.98 (0.03) [92.6]
	$\tau^2$	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.02 (0.03)
Separate Longitudinal Model						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.00 (0.03) [95.1]	2.00 (0.04) [94.4]	2.01 (0.04) [92.6]	2.02 (0.04) [92.1]	2.03 (0.04) [87.4]
	Random MA $\tau^2$	2.00 (0.03) [96.5]	2.00 (0.04) [95.7]	2.01 (0.04) [94.4]	2.02 (0.04) [93.2]	2.03 (0.04) [90.2]
Separate Time-to-event Model						
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.98 (0.15) [93.3]	2.37 (0.13) [0.2]	1.79 (0.11) [0.0]	1.49 (0.09) [0.0]	1.30 (0.09) [0.0]
	Random MA $\tau^2$	3.00 (0.15) [95.5]	2.39 (0.13) [2.7]	1.80 (0.11) [0.0]	1.50 (0.09) [0.0]	1.30 (0.09) [0.0]
	Random MA	0.10 (0.14)	0.12 (0.14)	0.08 (0.10)	0.06 (0.09)	0.06 (0.08)
	$\tau^2$					

Table 18: Two-stage simulation results for homogenous treatment effect between studies, low event rate data for joint model and separate longitudinal and time-to-event models. Stated as mean pooled estimate (standard error) [percentage coverage]. Note, the parameter values that the data was simulated under are identified in the table as “true”



## Two-Stage Simulation Results

### Homogenous treatment effect between studies, high event rate data

Model	Scenario					$\alpha^{(2)} = 1$
	$\alpha^{(2)} = 0$	$\alpha^{(2)} = 0.25$	$\alpha^{(2)} = 0.5$	$\alpha^{(2)} = 0.75$		
Joint longitudinal and time-to-event model (joiner, proportional association, shared random effects, unspecified baseline hazard)						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.00 (0.03) [94.8]	2.01 (0.03) [95.5]	2.01 (0.04) [92.2]	2.02 (0.04) [89.2]	2.03 (0.03) [89.1]
	Random MA $\tau^2$	2.00 (0.03) [96.6]	2.01 (0.03) [96.0]	2.01 (0.04) [94.6]	2.02 (0.04) [91.9]	2.03 (0.03) [91.7]
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.99 (0.08) [93.5]	2.99 (0.08) [94.3]	2.98 (0.07) [93.7]	2.98 (0.08) [92.7]	2.97 (0.08) [93.2]
	Random MA $\tau^2$	3.00 (0.08) [95.1]	2.99 (0.07) [96.0]	2.98 (0.07) [95.4]	2.98 (0.08) [94.4]	2.98 (0.08) [95.8]
Association estimate (true $\alpha^{(2)}$ as in column headers)	Fixed MA	0.05 (0.08)	0.05 (0.07)	0.05 (0.07)	0.05 (0.08)	0.06 (0.08)
	Random MA $\tau^2$	0.00 (0.01) [96.0]	0.25 (0.01) [94.4]	0.50 (0.01) [91.1]	0.74 (0.01) [89.7]	0.98 (0.02) [84.4]
Separate Longitudinal model	Fixed MA	0.00 (0.01) [96.9]	0.25 (0.01) [96.2]	0.50 (0.01) [93.8]	0.74 (0.01) [92.3]	0.98 (0.02) [88.5]
	Random MA $\tau^2$	0.004 (0.01)	0.004 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)
Separate Time-to-event model						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.00 (0.03) [95.0]	2.01 (0.03) [95.1]	2.02 (0.04) [90.4]	2.03 (0.04) [86.2]	2.04 (0.03) [82.0]
	Random MA $\tau^2$	2.00 (0.03) [96.4]	2.01 (0.03) [95.7]	2.02 (0.04) [93.0]	2.03 (0.04) [88.9]	2.04 (0.03) [85.8]
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)
	Random MA $\tau^2$	3.00 (0.08) [94.3]	2.01 (0.08) [0.0]	1.18 (0.06) [0.0]	0.85 (0.05) [0.0]	0.69 (0.05) [0.0]
Separate Time-to-event model	Fixed MA	3.01 (0.08) [95.3]	2.02 (0.08) [0.0]	1.19 (0.06) [0.0]	0.85 (0.05) [0.0]	0.69 (0.05) [0.0]
	Random MA $\tau^2$	0.05 (0.08)	0.10 (0.09)	0.06 (0.06)	0.04 (0.05)	0.04 (0.05)

Table 19: Two-stage simulation results for homogeneous treatment effect between studies, high event rate data for joint model and separate longitudinal and time-to-event models. Stated as mean pooled estimate (standard error) [percentage coverage]. Note, the parameter values that the data was simulated under are identified in the table as “true”

Table 19: Two-stage simulation results for homogenous treatment effect between studies, high event rate data for joint model and separate longitudinal and time-to-event models. Stated as mean pooled estimate (standard error) [percentage coverage]. Note, the parameter values that the data was simulated under are identified in the table as "true"

## Two-Stage Simulation Results

**Heterogeneous treatment effect between studies, low event rate data**

Model	Scenario					
	$\alpha^{(2)} = 0$	$\alpha^{(2)} = 0.25$	$\alpha^{(2)} = 0.5$	$\alpha^{(2)} = 0.75$	$\alpha^{(2)} = 1$	
Joint longitudinal and time-to-event model (joiner, proportional association, shared random effects, unspecified baseline hazard)						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.01 (0.23) [26.6]	2.00 (0.22) [26.0]	2.01 (0.23) [22.8]	2.01 (0.22) [24.9]	2.02 (0.24) [23.4]
	Random MA	2.01 (0.23) [87.3]	2.00 (0.22) [89.2]	2.01 (0.23) [86.6]	2.01 (0.22) [89.3]	2.02 (0.23) [86.6]
	$\tau^2$	0.47 (0.18)	0.46 (0.18)	0.47 (0.17)	0.47 (0.18)	0.47 (0.18)
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.76 (0.27) [58.5]	2.84 (0.26) [63.6]	2.89 (0.25) [66.8]	2.86 (0.26) [62.5]	2.84 (0.27) [56.9]
	Random MA	2.84 (0.24) [82.2]	2.92 (0.25) [86.8]	2.97 (0.25) [90.6]	2.94 (0.25) [88.0]	2.92 (0.27) [86.2]
	$\tau^2$	0.28 (0.25)	0.34 (0.25)	0.41 (0.25)	0.43 (0.25)	0.44 (0.25)
Association estimate (true $\alpha^{(2)}$ as in column headers)	Fixed MA	0.00 (0.01) [95.8]	0.25 (0.02) [95.1]	0.49 (0.02) [95.1]	0.74 (0.03) [93.9]	0.98 (0.03) [87.9]
	Random MA	0.00 (0.01) [96.7]	0.25 (0.02) [96.5]	0.50 (0.02) [96.2]	0.74 (0.03) [95.2]	0.98 (0.02) [91.0]
	$\tau^2$	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.02 (0.03)
Separate Longitudinal model						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.01 (0.23) [26.1]	2.00 (0.22) [25.6]	2.01 (0.23) [22.1]	2.01 (0.22) [24.7]	2.03 (0.23) [24.4]
	Random MA	2.01 (0.23) [87.0]	2.00 (0.22) [89.2]	2.01 (0.23) [86.8]	2.01 (0.22) [89.3]	2.03 (0.23) [86.6]
	$\tau^2$	0.47 (0.18)	0.46 (0.18)	0.47 (0.17)	0.47 (0.18)	0.47 (0.18)
Separate Time-to-event model						
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.86 (0.27) [64.3]	2.32 (0.23) [4.2]	1.78 (0.16) [0.0]	1.47 (0.14) [0.0]	1.28 (0.12) [0.0]
	Random MA	2.98 (0.26) [88.7]	2.40 (0.23) [25.8]	1.81 (0.17) [0.0]	1.49 (0.14) [0.0]	1.29 (0.12) [0.0]
	$\tau^2$	0.42 (0.26)	0.36 (0.21)	0.26 (0.17)	0.20 (0.15)	0.16 (0.13)

Table 20: Two-stage simulation results for heterogeneous treatment effect between studies, low event rate data for joint model and separate longitudinal and time-to-event models. Stated as mean pooled estimate (standard error) [percentage coverage]. Note, the parameter values that the data was simulated under are identified in the table as “true”

Two-Stage Simulation Results						
Heterogeneous treatment effect between studies, high event rate data						
Model	Scenario					
	$\alpha^{(2)} = 0$	$\alpha^{(2)} = 0.25$	$\alpha^{(2)} = 0.5$	$\alpha^{(2)} = 0.75$	$\alpha^{(2)} = 1$	
Joint longitudinal and time-to-event model (joiner, proportional association, shared random effects, unspecified baseline hazard)						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	1.99 (0.23) [20.6]	2.02 (0.23) [23.5]	2.02 (0.23) [22.0]	2.01 (0.23) [20.3]	2.02 (0.23) [21.7]
	Random MA $\tau^2$	2.00 (0.23) [88.2] 0.47 (0.17)	2.02 (0.23) [88.0] 0.47 (0.18)	2.02 (0.23) [85.9] 0.47 (0.18)	2.01 (0.23) [88.1] 0.47 (0.17)	2.02 (0.23) [87.2] 0.47 (0.18)
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.85 (0.24) [41.1]	2.89 (0.25) [40.0]	2.89 (0.25) [40.7]	2.90 (0.25) [42.6]	2.90 (0.25) [43.8]
	Random MA $\tau^2$	2.98 (0.23) [88.4] 0.45 (0.20)	2.98 (0.24) [88.0] 0.44 (0.20)	2.98 (0.24) [86.3] 0.46 (0.20)	2.97 (0.24) [87.2] 0.45 (0.20)	2.97 (0.25) [86.2] 0.45 (0.21)
Association estimate (true $\alpha^{(2)}$ as in column headers)	Fixed MA	0.00 (0.01) [93.9]	0.25 (0.01) [95.4]	0.50 (0.01) [91.8]	0.74 (0.01) [87.9]	0.98 (0.02) [83.6]
	Random MA $\tau^2$	0.00 (0.01) [96.0] 0.003 (0.01)	0.25 (0.01) [97.0] 0.005 (0.01)	0.50 (0.01) [94.1] 0.01 (0.01)	0.74 (0.01) [90.6] 0.01 (0.01)	0.98 (0.02) [87.9] 0.01 (0.02)
Separate Longitudinal model						
Longitudinal treatment effect coefficient (true $\beta_{12} = 2$ )	Fixed MA	2.00 (0.23) [22.1]	2.02 (0.23) [23.5]	2.03 (0.23) [21.8]	2.02 (0.23) [20.1]	2.03 (0.23) [22.6]
	Random MA $\tau^2$	2.00 (0.23) [88.6] 0.47 (0.17)	2.02 (0.23) [88.3] 0.47 (0.18)	2.03 (0.23) [85.7] 0.47 (0.18)	2.02 (0.23) [88.3] 0.47 (0.17)	2.03 (0.23) [87.1] 0.47 (0.18)
Separate Time-to-event model						
Time-to-event treatment effect coefficient (log(HR)) (true $\beta_{21} = 3$ )	Fixed MA	2.86 (0.24) [40.1]	1.97 (0.19) [0.0]	1.18 (0.11) [0.0]	0.85 (0.08) [0.0]	0.69 (0.07) [0.0]
	Random MA $\tau^2$	2.99 (0.23) [88.2] 0.45 (0.20)	2.03 (0.20) [0.9] 0.36 (0.15)	1.19 (0.11) [0.0] 0.19 (0.10)	0.85 (0.08) [0.0] 0.12 (0.08)	0.69 (0.07) [0.0] 0.09 (0.08)
Table 21: Two-stage simulation results for heterogeneous treatment effect between studies, high event rate data for joint model and separate longitudinal and time-to-event models. Stated as mean pooled estimate (standard error) [percentage coverage]. Note, the parameter values that the data was simulated under are identified in the table as "true"						

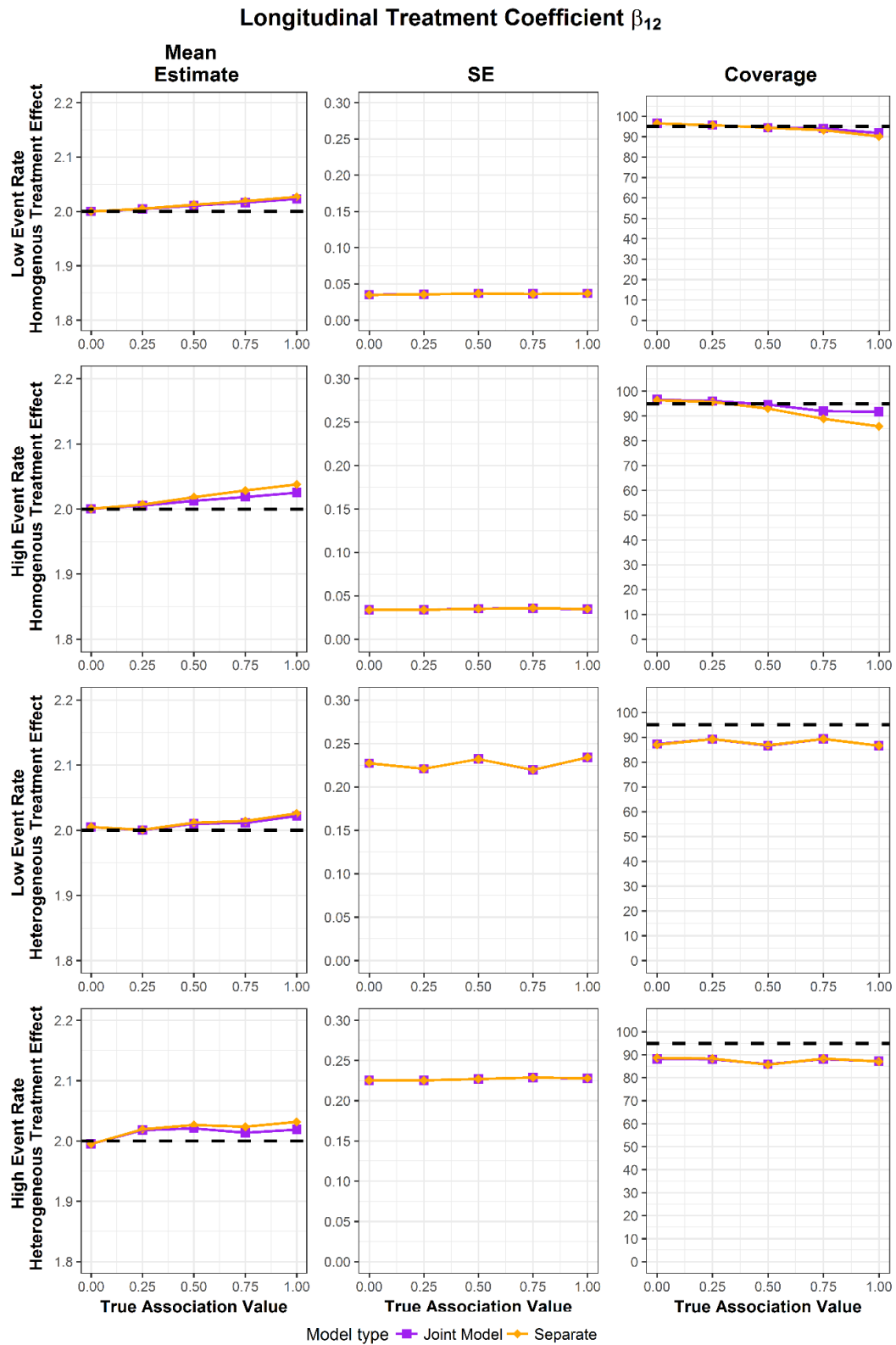


Figure 22: Plot for two-stage meta-analysis simulation study for longitudinal treatment effect parameter  $\beta_{12}$ . The dotted line in the mean estimate column identifies the “true” value the parameter was simulated under. The dotted line in the coverage column identifies 95% coverage.

### Survival Treatment Coefficient $\beta_{21}$

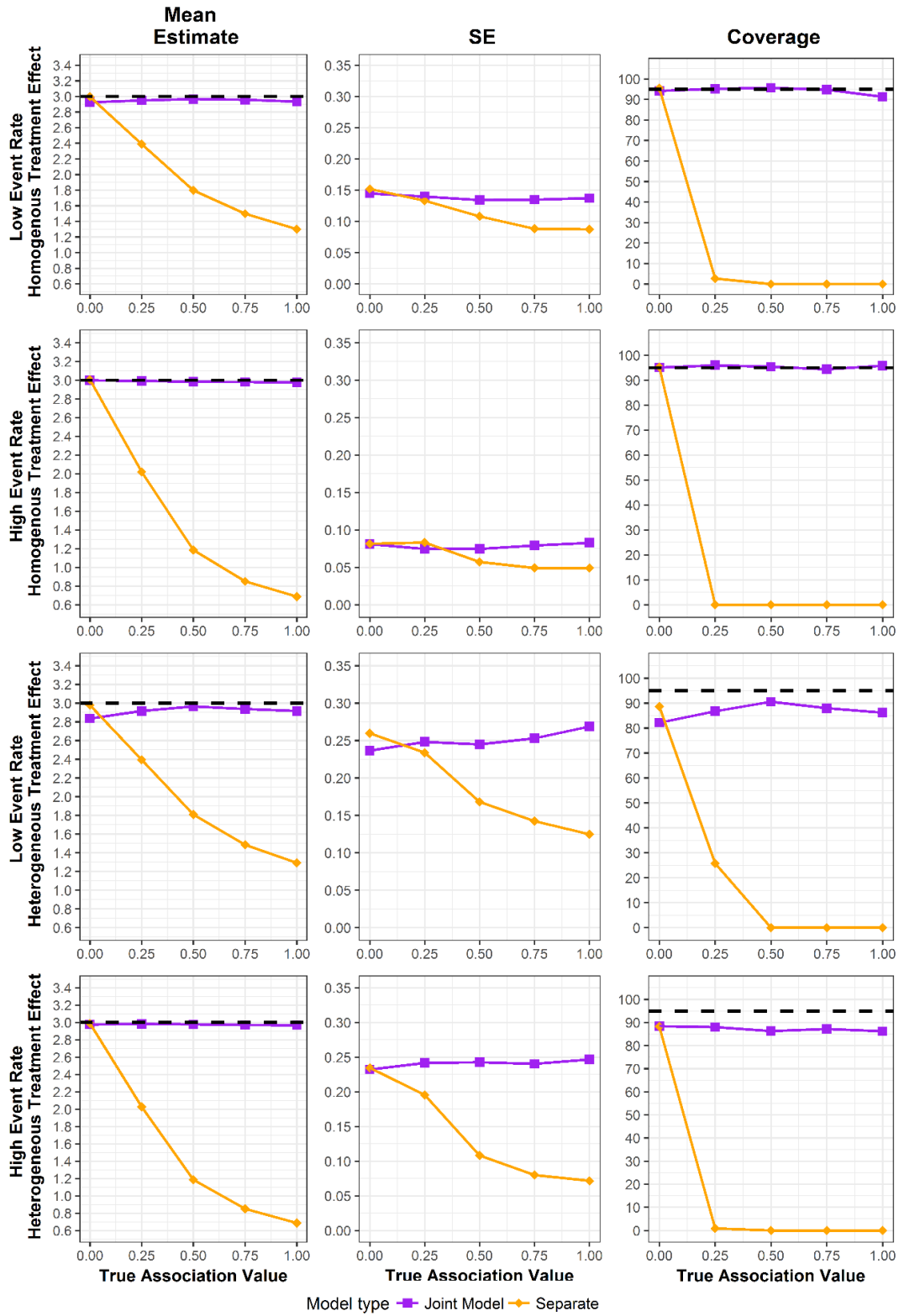


Figure 23: Plot for two-stage meta-analysis simulation study for time-to-event treatment effect parameter  $\beta_{21}$ . The dotted line in the mean estimate column identifies the “true” value the parameter was simulated under. The dotted line in the coverage column identifies 95% coverage.

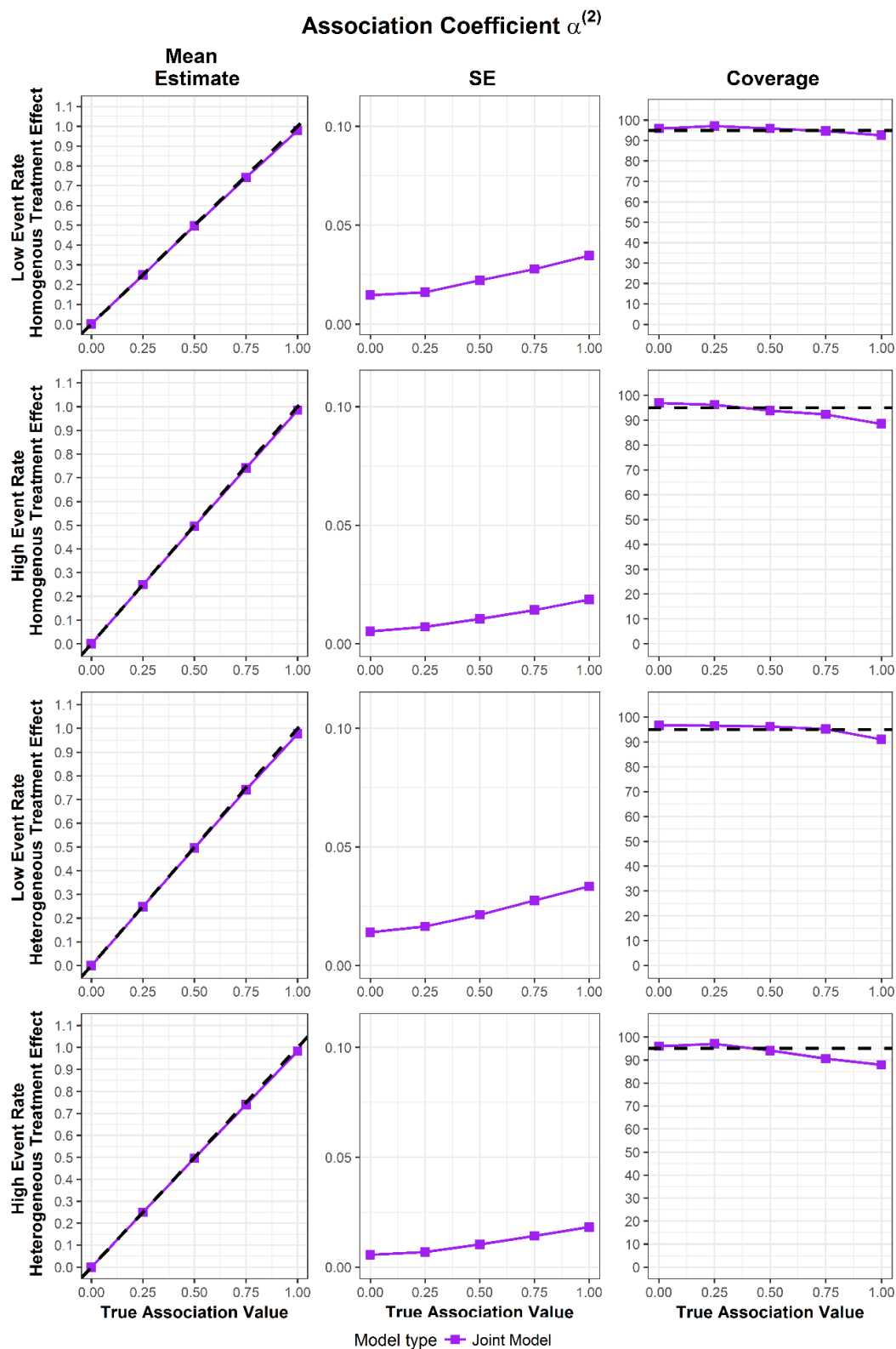


Figure 24: Plot for two-stage meta-analysis simulation study for association parameter  $\alpha^{(2)}$ . The dotted line in the mean estimate column identifies the “true” value the parameter was simulated under. The dotted line in the coverage column identifies 95% coverage.

## 7.2 Simulations investigating one-stage methods

A variety of ways exist to account for between study heterogeneity in one-stage meta-analytic models, including fixed interaction terms with study membership variables, study level random effects and stratification of baseline hazard by study. However, the reliability of these methods may depend a range of factors such as the number of studies included in the meta-analysis, the level of between study heterogeneity, or the level of association between the longitudinal and time-to-event outcomes.

Consequently, three sets of simulation scenarios were investigated to test the behaviour of the methods developed in Section 3.3. The first varied levels of association at both the individual and the study level, the second varied the numbers of included studies in the meta-analysis, and the third varied the level of between study heterogeneity.

As with the two-stage simulation study, it is interesting to compare the performance of one-stage joint meta-analytic models, with one-stage separate longitudinal or time-to-event meta-analytic models. As such, both approaches (separate and joint) have been conducted and compared.

In the two-stage simulation study, heterogeneity was induced between studies by generating treatment effect as realisations from a distribution. However, it was necessary during the one-stage simulations to exactly specify the distribution of the study level random effects, to allow the between study heterogeneity to be more accurately controlled. As such, I wrote and implemented a function to simulate multi-study joint longitudinal and time-to-event data (see Sections 4.4.1 and 5.2).

### 7.2.1 Simulation of joint data for one-stage simulations

Across the sets of scenarios investigated in the simulation study, multi-study joint data was generated under the same process using the methodology described previously in Sections 4.4.1 and 5.2. This methodology extends the work of Bender et al [219] and Austin [220]. In this section, this process is described, and details of parameters for each set of simulations are given. For each set of simulations, for each scenario, 1000 datasets were simulated.

The number of included studies varies between scenarios, however each simulated study contained 500 individuals randomised equally to two treatment groups. A maximum of 10 longitudinal measurements at times 0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 were permitted, with measurements recorded only up to the individual's survival time. Data for all studies

was simulated simultaneously, with any between study heterogeneity generated through specification of the study level random effects distribution. The joint data was simulated under the following model:

$$\begin{aligned} Y_{kij} &= \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij} \\ \lambda_{ki}(t) &= \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t)) \\ W_{2ki}(t) &= \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki}) \end{aligned} \quad (71)$$

In equation (71), the longitudinal outcome  $Y_{kij}$  follows a linear mixed effects model containing fixed intercept, time and treatment terms (with coefficients  $\beta_{10}$ ,  $\beta_{11}$  and  $\beta_{12}$ ), individual level random intercept and time terms ( $b_{0ki}^{(2)}$  and  $b_{1ki}^{(2)}$ ), study level random intercept and treatment effect terms ( $b_{0k}^{(3)}$  and  $b_{1k}^{(3)}$ ) and an error term  $\varepsilon_{kij}$ . The random effects follow multivariate normal distributions, with the individual level random effects distributed  $\mathbf{b}_{ki}^{(2)} \sim N(0, \mathbf{D})$ , and the study level random effects distributed  $\mathbf{b}_k^{(3)} \sim N(0, \mathbf{A})$ . The random effects are independent of each other, and of the error terms, which are considered to be IID following  $\varepsilon_{kij} \sim N(0, \sigma_e^2)$ .

The event times are modelled using a Cox PH model with unspecified baseline hazard  $\lambda_0(t)$ , with a single fixed effect of treatment group (coefficient  $\beta_{21}$ ). The longitudinal and time-to-event sub-models are linked through shared zero mean random effects (represented by term  $W_{2ki}(t)$ ), with common association parameters at each level of the random effects.

The event times  $\mathbf{T}_E$  were specified to be Gompertz distributed with mean 3 and standard deviation 0.5. Using the extreme value distribution (with  $\gamma \approx 0.5772$  representing Euler's constant, and  $\pi \approx 3.142$  the mathematical constant pi), this lead to the parameters controlling the event times distributions to be set to:

$$\begin{aligned} \theta_1 &= \frac{\pi}{\sqrt{6}\sigma_0} = \frac{\pi}{(0.5)\sqrt{6}} \approx 2.5651 \\ \theta_0 &= \log(\theta_1 \exp(-\gamma - \mu_0\theta_1)) = \log(\theta_1 \exp(-\gamma - 3\theta_1)) \approx -7.330517 \end{aligned}$$

Due to the volume of planned simulations, only datasets with a “low” (~25%) event rate were generated. A range of censoring parameters were tested in the multi-study data simulation function to obtain datasets with mean event rate at 25%. As such, the censoring times  $\mathbf{T}_C$  followed an exponential distribution with parameter  $\varphi = \exp(-0.426)$ . As before, the survival time for each individual  $i$  was the minimum of their censoring and event times ( $T_{Ski} = \min(T_{Eki}, T_{Cki})$ ).



As mentioned, when the simulation studies were designed, parameter values were chosen such that deviations of different methods from the true parameters values would be clearly discernible. A summary of the values used for the different groups of simulations is given in Table 22.

Briefly, all sets of simulations utilised the same fixed effect and error term variance values ( $\beta_{10} = 1, \beta_{11} = 3, \beta_{12} = 2, \beta_{21} = 3, \sigma_e^2 = 0.01$ ). Additionally, the individual level random effects covariance matrix  $\mathbf{D}$  remained constant (defined in Table 22). However, the remaining aspects of the datasets (association parameters, number of included studies, level of between study heterogeneity) varied between simulation groups. These aspects are stated in Table 22 for each simulation group, and are briefly discussed in the following section.

#### 7.2.1.1 Set 1: Varying levels of association

The individual level association parameter  $\alpha^{(2)}$  and the study level association parameter  $\alpha^{(3)}$  are permitted to take values 0, 0.5 and 1, giving a total of 9 unique scenarios (all combinations of possible association parameters at each level). Only positive associations were examined as situations with negative associations were expected to have similar behaviour, but in the opposite direction. The number of included studies in each dataset was set to 5, whilst the study level random effects covariance matrix  $\mathbf{A}$  (defined in Table 22) remained constant across scenarios.

#### 7.2.1.2 Set 2: Varying numbers of studies included in the meta-analysis

During this set of simulations, meta-analyses containing 5, 10 or 15 studies were generated. These values were selected to give a sense of the effect a range of different numbers of included studies on the behaviour of the developed methods. The association parameters were held constant across scenarios (with  $\alpha^{(2)} = \alpha^{(3)} = 0.5$ ). Additionally, the study level random effects covariance matrix  $\mathbf{A}$  (defined in Table 22) remained constant across scenarios.

#### 7.2.1.3 Set 3: Varying levels of between study heterogeneity

Between study heterogeneity was controlled through the study level random effects covariance matrix  $\mathbf{A}$ . Values taken for  $\mathbf{A}$ , labelled  $\mathbf{A}_1$ ,  $\mathbf{A}_2$  and  $\mathbf{A}_3$  are specified in Table 22, representing cases for no between study heterogeneity, and then two increasing levels of between study heterogeneity.

During this simulation set, across all scenarios, 5 studies were simulated for each dataset, with association parameters were held constant across scenarios at  $\alpha^{(2)} = \alpha^{(3)} = 0.5$ .

	Set 1: Varying association parameters	Set 2: Varying number of included studies	Set 3: Varying level of between study heterogeneity
Number of included studies	5	5, 10, 15	5
Number of individuals within each study	500	500	500
Measurement times	0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4	0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4	0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4
Longitudinal fixed effect parameters ( $\beta_{10}, \beta_{11}, \beta_{12}$ )	$\beta_{10} = 1, \beta_{11} = 3, \beta_{12} = 2$	$\beta_{10} = 1, \beta_{11} = 3, \beta_{12} = 2$	$\beta_{10} = 1, \beta_{11} = 3, \beta_{12} = 2$
Time-to-event fixed effect parameters ( $\beta_{21}$ )	$\beta_{21} = 3$	$\beta_{21} = 3$	$\beta_{21} = 3$
Individual level association parameter ( $\alpha^{(2)}$ )	$\alpha^{(2)} = (0, 0.5, 1)$	$\alpha^{(2)} = 0.5$	$\alpha^{(2)} = 0.5$
Individual level random effects covariance matrix ( $D$ )	$D = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$	$D = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$	$D = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$
Study level association parameter ( $\alpha^{(3)}$ )	$\alpha^{(3)} = (0, 0.5, 1)$	$\alpha^{(3)} = 0.5$	$\alpha^{(3)} = 0.5$
Study level random effects covariance matrix ( $A$ )	$A = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$	$A = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$	$A_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ $A_2 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}$ $A_3 = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$
Error term variance ( $\sigma_e^2$ )	0.01	0.01	0.01
Parameters controlling event time distribution ( $\theta_0, \theta_1$ )	$\theta_1 = \frac{\pi}{(0.5)\sqrt{6}}$ $\theta_0 = \log(\theta_1 \exp(-\gamma - 3\theta_1))$	$\theta_1 = \frac{\pi}{(0.5)\sqrt{6}}$ $\theta_0 = \log(\theta_1 \exp(-\gamma - 3\theta_1))$	$\theta_1 = \frac{\pi}{(0.5)\sqrt{6}}$ $\theta_0 = \log(\theta_1 \exp(-\gamma - 3\theta_1))$
Parameter controlling censoring time distribution ( $\varphi$ )	$\exp(-0.426)$	$\exp(-0.426)$	$\exp(-0.426)$

Table 22: Parameters used for simulation of one-stage datasets under different simulation groups (note that  $\gamma \approx 0.5772$  representing Euler's constant, and  $\pi \approx 3.142$ ). The `simjointmeta()` function in the `jointmeta` package was used to simulate the data.

### 7.2.2 One-stage models fitted to simulated joint data

The models fitted to each set of simulations in the one-stage simulation study were described in Section 3.3.2. Exact specifications of the joint models are given in Table 23.

Briefly, the examined models fall into 6 unique groups (identified in Table 23). Group 0 is a naïve model which ignores between study heterogeneity. Group 1 accounts for between study heterogeneity using interaction terms between the study membership and treatment group variables in both the longitudinal and time-to-event sub-models. Group 2 introduces a study level random treatment effect which is shared alongside the individual level random effects between sub-models, but the sub-models still contain a fixed study membership term. Group 3 solely accounts for between study heterogeneity using a study level random intercept and treatment effect, which are shared between sub-models through the association structure. Group 4 includes fixed interaction terms between study membership and treatment group variables in the longitudinal sub-model, and a baseline hazard stratified by study in the time-to-event sub-model. Group 5 includes a study membership fixed effect in the longitudinal sub-model, a baseline hazard stratified by study in the time-to-event sub-model, and a study level random treatment effect present in both sub-models through the association structure.

As well as one-stage joint models, separate longitudinal and time-to-event one-stage models were also fitted to each dataset in each scenario in each simulation group. The longitudinal separate models had the same specification as the corresponding joint model longitudinal sub-model within each model group. The time-to-event separate models had the same specification as the corresponding joint model time-to-event sub-model within each model group, apart from the absence of the  $W_{2ki}(t)$  term.

Once each model had been fitted, a record was made of whether there was an error with the model fit, and whether the model had converged. Separate models were fitted automatically during the joint modelling process, to obtain starting values for the EM algorithm (Section 4.3.2.1); consequently the number of failed fits was equal between the separate and joint approaches. If the model had successfully fitted and converged, any parameters of interest were extracted along with their 95% confidence intervals. These included any longitudinal or time-to-event treatment parameters ( $\beta_{12}$  and  $\beta_{21}$  terms), and association parameters ( $\alpha^{(2)}$  and if present  $\alpha^{(3)}$ ).

For models groups that involved interaction terms between study membership and treatment group (groups 1 and 4), study specific treatment effects could be extracted. In

order to be able to compare across studies, these study specific treatment effects were pooled using methods identical to a random effects meta-analysis (described in Section 3.2.3) to produce a single estimate for treatment effect.

When the information from each model fit was extracted, a record was made of the number of successful model fits for each scenario within each simulation group.

Additionally, for each of the joint and separate models, for each parameter of interest, the mean parameter estimate and its standard error (standard deviation of the parameter estimates) were calculated. In addition, the coverage was estimated (the proportion of the 1000 model fits whose estimated 95% confidence interval for the parameter in question included the true parameter value).

Model Group	Model component	Equation
0	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
1	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + \beta_{14}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + \beta_{23}treat_{ki} * study_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
2	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + \beta_{22}study_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$
3	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_0(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{0k}^{(3)} + b_{1k}^{(3)}treat_{ki})$
4	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + \beta_{14}treat_{ki} * study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}t_{kij} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski})$
5	Longitudinal Sub-Model	$Y_{kij} = \beta_{10} + \beta_{11}t_{kij} + \beta_{12}treat_{ki} + \beta_{13}study_{ki} + b_{0ki}^{(2)} + b_{1ki}^{(2)}time_{kij} + b_{1k}^{(3)}treat_{ki} + \varepsilon_{kij}$
	Time-to-event Sub-Model	$\lambda_{ki}(t) = \lambda_{0k}(t) \exp(\beta_{21}treat_{ki} + W_{2ki}(t))$
	Association Structure	$W_{2ki}(t) = \alpha^{(2)}(b_{0ki}^{(2)} + b_{1ki}^{(2)}T_{Ski}) + \alpha^{(3)}(b_{1k}^{(3)}treat_{ki})$

Table 23: Model group specifications for one-stage simulation studies

### 7.2.3 One-stage simulation study results

Throughout, mean parameter estimates along with their standard errors and coverage values are presented in tables. Two sets of figures were produced for each simulation group, the first is simply a graphical representation of the mean parameter estimate with 95% confidence intervals calculated based on the SE of the parameter estimates within each scenario (included in the main text). The second set plots the parameter point estimates and confidence intervals for each successful model fit within each scenario, allowing an examination of the widths of confidence intervals, and the positioning of estimates compared to the true value of the parameter. These graphs are numerous, and so are presented in Appendix 4.

#### 7.2.3.1 *Simulation Set 1: Varying levels of association*

For the first set of simulations (Table 24 and Table 25, and Figure 25-Figure 28), the lowest proportion of successfully completed model fits was 94.2% (for model group 1 fitted to the scenario where  $\alpha^{(2)} = 0$  and  $\alpha^{(3)} = 1$ ).

The mean estimate for the longitudinal treatment effect parameter  $\beta_{12}$  is close to the true value of 2 for both the separate and the joint analyses across all model groups and scenarios (Table 24 and Table 25, Figure 25). Additionally the SE of the point estimates for the treatment effect produced by the simulations is similar for both the separate and joint analyses. However, there are noticeable differences in coverage. For both the separate and joint analyses, for model group 0 (the naïve model that ignores between study heterogeneity), coverage is under 20% for all scenarios investigated (Table 24, Figure 25). However, for the remaining model groups (which all account for between study heterogeneity in various ways), the coverage for the separate longitudinal model is higher, between 85 and 90% across all scenarios. Whilst this holds for the joint model results for model groups 1 and 4, the same cannot be said for the joint model results from any model group involving study level random effects (model groups 2, 3 and 5). Here, although the mean parameter estimate is close to the true value, the coverage is low (between 9 and 14%). An explanation for this can be found by examining the plots showing the point estimate and confidence interval for each dataset for each model group and scenario (Appendix 4, Figure 40-Figure 42). For any model group containing study level random effects, the confidence intervals for the joint model estimates of longitudinal treatment effect are narrow. Consequently, although the estimates are clustered around the true parameter value, coverage is low.

For the time-to-event treatment effect parameter  $\beta_{21}$  (Table 24 and Table 25, Figure 26), as study level association ( $\alpha^{(3)}$ ) increases, for both separate and joint approaches, the variability between estimates increases (represented by larger confidence bands about the mean estimates). Additionally, as  $\alpha^{(3)}$  increases, the estimates from model groups 0 and 3 become increasingly different from the estimates from the other model groups. As individual level association ( $\alpha^{(2)}$ ) increases, the separate time-to-event approach increasingly underestimates the treatment effect, whilst the estimates from the joint approach remain close to the true value. The coverage of the results from the joint models is comparable to that of the separate models when the individual level association  $\alpha^{(2)} = 0$ , however the joint modelling results show better (although not ideal) coverage for non-zero individual level associations (Table 24 and Table 25, and Figure 43-Figure 45 in Appendix 4). Non-zero study level association appears to increase the variability between simulations in estimated mean time-to-event treatment effect, rather than increase the discrepancy between the separate and joint modelling results. Throughout, with non-zero association at any level, naïve model group 0 performs badly, for both separate and joint approaches.

All fitted joint models contain individual level random effects, and so all estimated the individual level association parameter  $\alpha^{(2)}$  (note: no separate analyses produce estimates of either association parameter). The estimates of  $\alpha^{(2)}$  appear poor for model group 0 (which ignores between study heterogeneity). However, for the remaining model groups, across all investigated scenarios, the mean association parameter estimates are close to the true values (Table 24 and Table 25, Figure 27). The coverage appears constant and high for model groups 1, 2, 4 and 5, however there are some instances of lower coverage for model group 3 (which solely relies on study level random effects to quantify between study heterogeneity). Examination of Figure 46-Figure 48 in Appendix 4 (that show the results of the individual simulations), shows a tendency of underestimation of the individual level association parameter in model group 3 for combined higher levels of individual and study level association.

The study level association parameter  $\alpha^{(3)}$  is only estimated for model groups that contain study level random effects (namely model groups 2 and 5 which contain study level random treatment effect, and model group 3 which contains study level random intercept and treatment effect). The mean estimates of  $\alpha^{(3)}$  are closest to the true values for model group 3, although they are still biased (Table 24 and Table 25, Figure 28). However, the



estimates for model groups 2 and 5 are not close to the true values. The coverage appears varied for the study level association parameter across all model groups, with coverage higher for model group 3. From Figure 49-Figure 51 in Appendix 4, the confidence intervals for the distinct simulations can be seen to be wide, especially for model groups 2 and 5.

	Association Parameters		Number of successful joint fits	Longitudinal Treatment Effect ( $\beta_{12} = 2$ )		Time-to-event treatment effect ( $\beta_{21} = 3$ )		Association Parameters	
	True Value	$\alpha^{(3)}$		Separate model	Joint Model	Separate Model	Joint Model	Joint Model $\alpha^{(2)}$	Joint Model $\alpha^{(3)}$
Group 0	0	0	1000	2.02 (0.53) [18.6]	2.02 (0.53) [17.9]	3.02 (0.15) [95.4]	3.02 (0.15) [94.7]	0.00 (0.01) [93.4]	NA
	0	0.5	1000	2.00 (0.56) [17.6]	2.00 (0.56) [16.6]	2.50 (0.35) [24.5]	2.57 (0.32) [31.1]	0.06 (0.04) [18.6]	NA
	0	1	1000	2.00 (0.55) [17.0]	2.00 (0.55) [15.9]	1.93 (0.48) [5.6]	2.04 (0.46) [7.5]	0.10 (0.06) [12.2]	NA
	0.5	0	1000	2.00 (0.56) [17.2]	2.00 (0.56) [16.6]	1.63 (0.11) [0.0]	2.46 (0.26) [14.8]	0.36 (0.07) [5.5]	NA
	0.5	0.5	1000	2.01 (0.55) [18.5]	2.01 (0.55) [17.0]	1.57 (0.18) [0.0]	2.74 (0.29) [43.5]	0.46 (0.04) [49.7]	NA
	0.5	1	1000	1.97 (0.55) [18.9]	1.97 (0.55) [18.0]	1.45 (0.25) [0.0]	2.35 (0.46) [15.5]	0.48 (0.06) [45.9]	NA
	1	0	1000	1.99 (0.55) [18.7]	1.99 (0.55) [17.5]	1.11 (0.09) [0.0]	1.84 (0.38) [2.0]	0.57 (0.15) [0.5]	NA
	1	0.5	1000	2.04 (0.55) [19.7]	2.04 (0.55) [18.5]	1.12 (0.14) [0.0]	2.29 (0.42) [13.2]	0.76 (0.11) [4.7]	NA
	1	1	1000	1.97 (0.55) [18.0]	1.97 (0.55) [16.9]	1.13 (0.21) [0.0]	2.34 (0.55) [18.3]	0.84 (0.12) [22.1]	NA
	0	0	1000	2.02 (0.53) [87.1]	2.02 (0.53) [89.3]	3.04 (0.20) [75.0]	3.04 (0.15) [85.5]	0.00 (0.01) [93.8]	NA
Group 1	0	0.5	1000	2.00 (0.55) [87.1]	2.00 (0.55) [88.2]	3.02 (0.54) [57.3]	3.05 (0.27) [84.3]	0.00 (0.01) [93.0]	NA
	0	1	942	2.01 (0.55) [87.9]	2.01 (0.55) [87.9]	2.98 (0.98) [55.1]	3.05 (0.49) [82.9]	0.00 (0.01) [95.2]	NA
	0.5	0	1000	2.00 (0.56) [86.4]	2.00 (0.56) [87.5]	1.64 (0.17) [0.0]	3.03 (0.14) [86.3]	0.51 (0.02) [92.6]	NA
	0.5	0.5	1000	2.01 (0.55) [86.2]	2.01 (0.55) [87.5]	1.69 (0.47) [6.0]	3.04 (0.26) [85.0]	0.51 (0.02) [92.8]	NA
	0.5	1	977	1.97 (0.55) [86.7]	1.97 (0.55) [87.3]	1.79 (0.86) [27.1]	3.02 (0.46) [87.1]	0.51 (0.02) [88.3]	NA
	1	0	1000	1.99 (0.55) [88.8]	1.99 (0.55) [89.3]	1.12 (0.15) [0.0]	3.04 (0.14) [86.7]	1.02 (0.03) [87.8]	NA
	1	0.5	1000	2.04 (0.54) [87.4]	2.04 (0.54) [87.9]	1.12 (0.33) [0.1]	3.06 (0.28) [82.3]	1.02 (0.03) [83.2]	NA
	1	1	998	1.97 (0.54) [87.2]	1.97 (0.54) [87.7]	1.16 (0.59) [3.3]	3.09 (0.52) [82.9]	1.03 (0.04) [75.6]	NA
	0	0	1000	2.02 (0.53) [89.4]	2.02 (0.53) [10.1]	3.03 (0.15) [95.4]	3.04 (0.15) [93.7]	0.00 (0.01) [94.1]	-0.002 (0.24) [98.7]
	0	0.5	1000	2.00 (0.55) [88.3]	2.00 (0.55) [10.1]	3.11 (0.28) [65.1]	3.11 (0.28) [64.7]	0.00 (0.01) [94.2]	0.039 (0.23) [38.1]
Group 2	0	1	1000	2.01 (0.55) [87.8]	2.01 (0.55) [9.7]	3.26 (0.51) [32.4]	3.26 (0.51) [33.5]	0.00 (0.02) [94.0]	0.067 (0.26) [8.5]
	0.5	0	999	2.00 (0.56) [87.3]	2.00 (0.56) [10.0]	1.64 (0.11) [0.0]	3.03 (0.14) [93.9]	0.50 (0.02) [94.9]	0.003 (0.24) [99.1]
	0.5	0.5	1000	2.01 (0.55) [87.3]	2.01 (0.55) [11.0]	1.76 (0.21) [0.0]	3.10 (0.27) [62.5]	0.51 (0.02) [93.7]	0.045 (0.55) [32.3]
	0.5	1	1000	1.98 (0.55) [87.9]	1.98 (0.55) [12.0]	2.01 (0.42) [3.4]	3.22 (0.47) [39.3]	0.51 (0.03) [86.7]	0.072 (0.27) [7.8]
	1	0	1000	1.99 (0.55) [89.3]	1.99 (0.55) [11.3]	1.11 (0.09) [0.0]	3.05 (0.14) [92.7]	1.02 (0.03) [90.6]	-0.012 (0.32) [99.2]
	1	0.5	1000	2.04 (0.54) [87.8]	2.04 (0.54) [12.0]	1.18 (0.15) [0.0]	3.10 (0.28) [62.5]	1.01 (0.04) [91.6]	0.026 (0.30) [28.6]
	1	1	999	1.97 (0.54) [87.6]	1.97 (0.54) [10.7]	1.35 (0.30) [0.0]	3.24 (0.52) [35.7]	1.01 (0.04) [88.2]	0.049 (0.24) [5.4]
	0	0	1000	2.02 (0.53) [89.4]	2.02 (0.53) [10.1]	3.03 (0.15) [95.4]	3.04 (0.15) [93.7]	0.00 (0.01) [94.1]	-0.002 (0.24) [98.7]
	0	0.5	1000	2.00 (0.55) [88.3]	2.00 (0.55) [10.1]	3.11 (0.28) [65.1]	3.11 (0.28) [64.7]	0.00 (0.01) [94.2]	0.039 (0.23) [38.1]
	0	1	1000	2.01 (0.55) [87.8]	2.01 (0.55) [9.7]	3.26 (0.51) [32.4]	3.26 (0.51) [33.5]	0.00 (0.02) [94.0]	0.067 (0.26) [8.5]

Table 24: Results for simulations for longitudinal treatment effect  $\beta_{12}$ , time-to-event treatment effect  $\beta_{21}$ , and association parameters ( $\alpha^{(2)}$  and if estimated  $\alpha^{(3)}$ ) for varying associations.. Results stated as mean parameter estimate (SE between simulation estimates) [coverage] for model groups 0-2. Model group 1 reports overall longitudinal and time-to-event treatment effect estimates produced by combining study specific treatment effect estimates (see Section 3.2.3). "True" parameter values are stated in the column headers. Where listed as NA, the parameter  $\alpha^{(3)}$  was not estimated by the model

	Association Parameters		Longitudinal Treatment Effect ( $\beta_{12} = 2$ )		Time-to-event treatment effect ( $\beta_{21} = 3$ )		Association Parameters	
	True Value	Number of successful joint fits	Separate model	Joint Model	Separate Model	Joint Model	Joint Model $\alpha^{(2)}$	Joint Model $\alpha^{(3)}$
Group 3	0	1000	2.02 (0.53) [89.5]	2.02 (0.53) [11.0]	3.02 (0.15) [95.4]	3.03 (0.15) [94.1]	0.00 (0.01) [94.1]	-0.002 (0.06) [96.4]
	0	1000	2.00 (0.55) [88.3]	2.00 (0.56) [11.2]	2.50 (0.35) [24.5]	2.90 (0.28) [67.6]	0.00 (0.01) [94.5]	0.429 (0.17) [44.8]
	0	1000	2.01 (0.55) [87.9]	2.01 (0.55) [10.3]	1.93 (0.48) [5.6]	2.67 (0.49) [37.2]	0.00 (0.02) [97.3]	0.753 (0.32) [24.7]
	0.5	1000	2.00 (0.56) [87.3]	2.00 (0.56) [10.7]	1.63 (0.11) [0.0]	3.02 (0.14) [94.8]	0.50 (0.02) [96.1]	-0.001 (0.06) [96.6]
	0.5	999	2.01 (0.55) [87.4]	2.01 (0.55) [11.4]	1.57 (0.18) [0.0]	2.88 (0.27) [62.9]	0.48 (0.03) [83.2]	0.427 (0.17) [42.8]
	0.5	1000	1.98 (0.55) [87.9]	1.98 (0.55) [12.7]	1.45 (0.25) [0.0]	2.64 (0.44) [35.6]	0.44 (0.04) [45.3]	0.744 (0.31) [23.2]
Group 4	1	1000	1.99 (0.55) [89.3]	1.99 (0.55) [11.6]	1.11 (0.09) [0.0]	3.04 (0.14) [92.7]	1.01 (0.03) [92.1]	-0.001 (0.05) [96.7]
	1	1000	2.04 (0.54) [87.8]	2.04 (0.54) [13.4]	1.12 (0.14) [0.0]	2.88 (0.29) [62.7]	0.97 (0.04) [79.8]	0.431 (0.17) [44.9]
	1	1000	1.97 (0.54) [87.5]	1.97 (0.54) [11.4]	1.13 (0.21) [0.0]	2.65 (0.49) [33.6]	0.88 (0.07) [31.1]	0.754 (0.32) [23.1]
	0	1000	2.01 (0.53) [87.1]	2.01 (0.53) [89.3]	3.02 (0.15) [95.8]	3.02 (0.15) [94.9]	0.00 (0.01) [94.2]	NA
	0	1000	2.00 (0.55) [87.1]	2.00 (0.55) [87.8]	3.03 (0.28) [72.5]	3.03 (0.28) [71.7]	0.00 (0.01) [94.3]	NA
	0	1000	2.01 (0.55) [87.9]	2.01 (0.55) [87.9]	3.08 (0.47) [44.8]	3.08 (0.47) [46.1]	0.00 (0.01) [96.2]	NA
Group 5	0.5	1000	2.00 (0.56) [86.4]	2.00 (0.56) [87.3]	1.66 (0.11) [0.0]	3.02 (0.14) [94.8]	0.50 (0.02) [96.1]	NA
	0.5	1000	2.01 (0.54) [86.5]	2.01 (0.54) [87.5]	1.73 (0.20) [0.0]	3.04 (0.26) [69.7]	0.50 (0.02) [95.2]	NA
	0.5	1000	1.98 (0.55) [86.9]	1.97 (0.55) [87.8]	1.87 (0.35) [1.1]	3.08 (0.43) [47.1]	0.50 (0.03) [91.7]	NA
	1	1000	1.99 (0.55) [88.8]	1.99 (0.55) [89.3]	1.13 (0.09) [0.0]	3.03 (0.14) [93.7]	1.01 (0.03) [93.0]	NA
	1	1000	2.04 (0.54) [87.4]	2.04 (0.54) [88.0]	1.17 (0.15) [0.0]	3.06 (0.28) [65.6]	1.01 (0.04) [93.8]	NA
	1	1000	1.98 (0.54) [87.5]	1.97 (0.54) [87.8]	1.29 (0.26) [0.0]	3.16 (0.49) [42.3]	1.00 (0.04) [88.8]	NA
Group 5	0	1000	2.00 (0.53) [89.8]	2.00 (0.53) [9.6]	3.02 (0.15) [95.8]	3.03 (0.15) [94.9]	0.00 (0.01) [94.6]	-0.010 (0.25) [99.0]
	0	1000	2.00 (0.55) [88.3]	2.00 (0.55) [9.5]	3.03 (0.28) [72.5]	3.04 (0.28) [73.2]	0.00 (0.01) [94.3]	0.045 (0.24) [45.8]
	0	1000	2.01 (0.55) [87.8]	2.01 (0.55) [9.5]	3.08 (0.47) [44.8]	3.08 (0.47) [46.0]	0.00 (0.01) [95.9]	0.067 (0.27) [12.2]
	0.5	999	2.00 (0.56) [87.3]	2.00 (0.56) [10.0]	1.66 (0.11) [0.0]	3.02 (0.14) [94.3]	0.50 (0.02) [95.5]	0.003 (0.26) [98.6]
	0.5	999	2.04 (0.58) [84.8]	2.04 (0.58) [10.6]	1.74 (0.20) [0.0]	3.04 (0.26) [70.2]	0.50 (0.02) [94.8]	0.030 (0.23) [41.3]
	0.5	1000	1.98 (0.55) [87.9]	1.98 (0.55) [11.6]	1.87 (0.35) [1.1]	3.09 (0.43) [48.3]	0.50 (0.03) [91.5]	0.071 (0.29) [10.3]
Group 5	1	1000	1.99 (0.55) [89.3]	1.99 (0.55) [11.1]	1.13 (0.09) [0.0]	3.04 (0.14) [93.6]	1.01 (0.04) [92.6]	-0.014 (0.39) [99.1]
	1	1000	2.04 (0.54) [87.8]	2.04 (0.54) [12.3]	1.17 (0.15) [0.0]	3.07 (0.28) [66.2]	1.01 (0.04) [93.0]	0.027 (0.36) [37.4]
	1	999	1.97 (0.54) [87.6]	1.97 (0.54) [10.4]	1.29 (0.26) [0.0]	3.16 (0.50) [41.6]	1.00 (0.04) [88.9]	0.047 (0.26) [8.2]

Table 25: Results for simulations for longitudinal treatment parameter  $\beta_{12}$ , time-to-event treatment parameter  $\beta_{21}$ , and association parameters ( $\alpha^{(2)}$  and if estimated  $\alpha^{(3)}$ ) for varying levels of association across different levels of the data. Results stated as mean parameter estimate (SE between simulation estimates) [coverage] for model groups 3-5. Model group 4 reports overall longitudinal treatment effect estimates produced by combining study specific treatment effect estimates (see section 7.2.2). True values for the association parameters are given in the 2nd and 3rd columns of the table. Note, NA identifies models where the  $\alpha^{(3)}$  was not estimated.

# Mean Longitudinal Treatment Estimate $\beta_{12}$ for varying association parameters

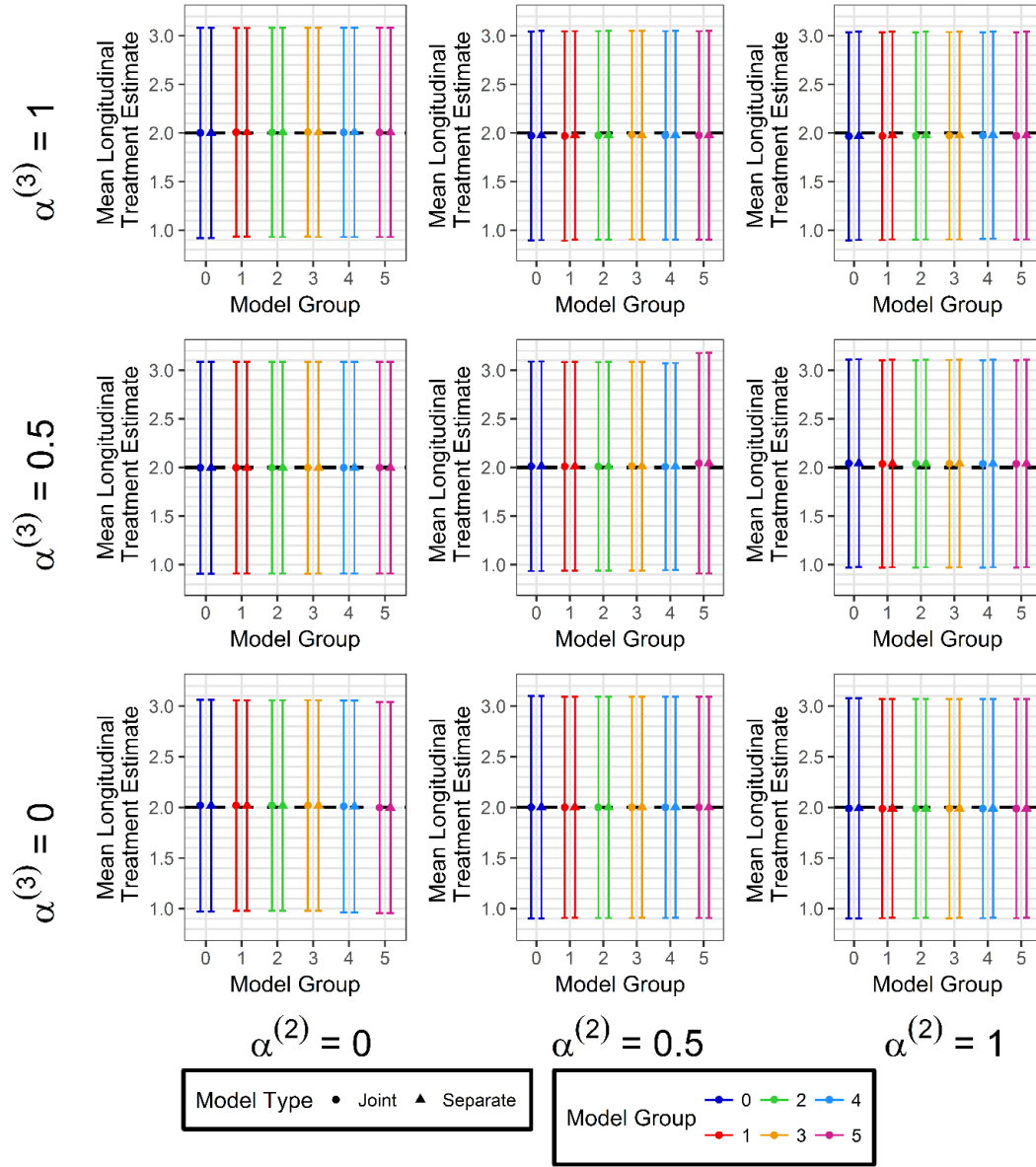


Figure 25: Graphical representation of the mean longitudinal treatment effect ( $\beta_{12}$ ) estimates from Table 24 and Table 25 for separate and joint models for simulation set 1, investigating varying association parameters. The dashed line identifies the "true" value of  $\beta_{12}$  that the data was simulated under.

# Mean Time-to-Event Treatment Estimate $\beta_{21}$ for varying association parameters

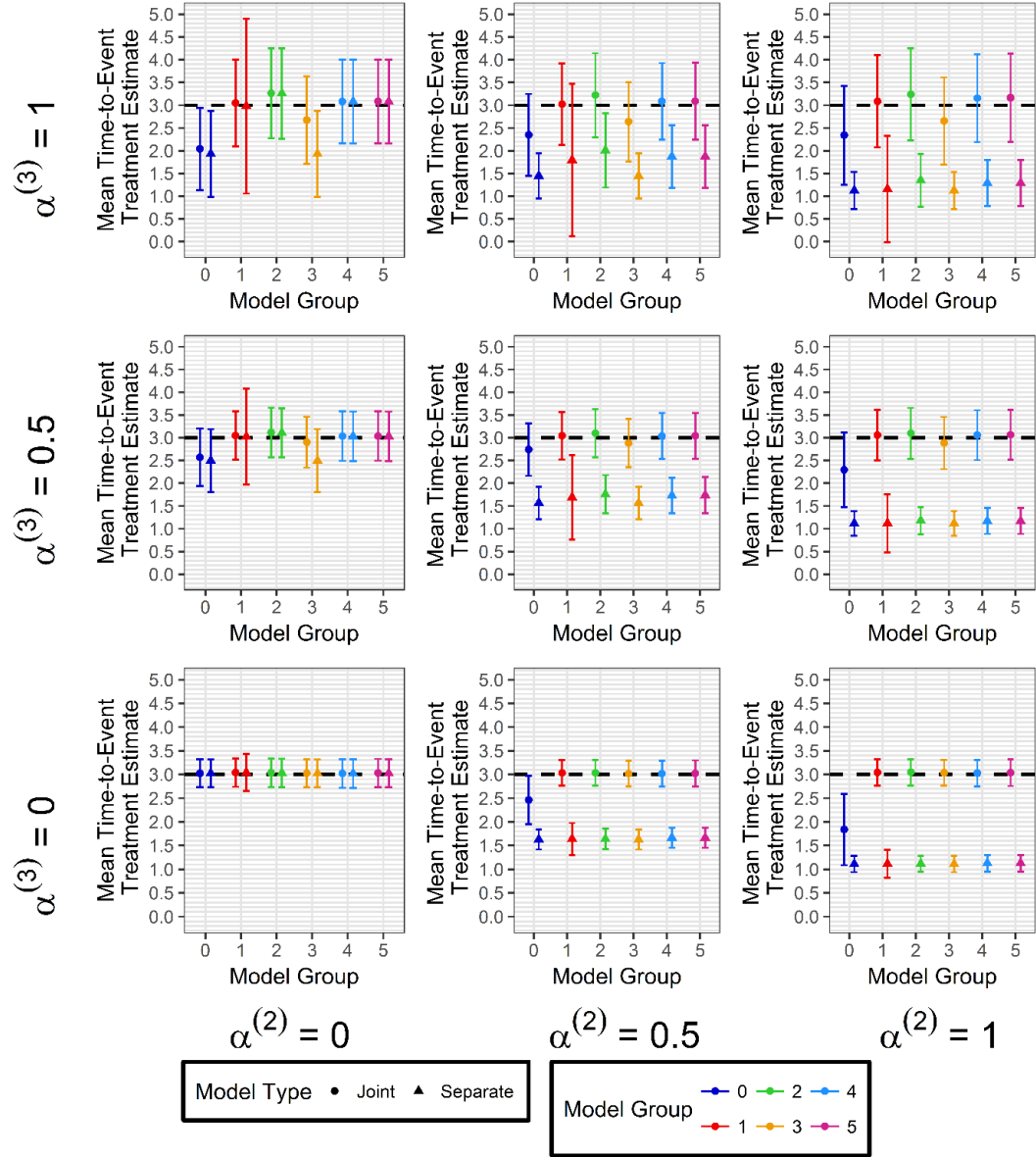


Figure 26: Graphical representation of the mean time-to-event treatment effect ( $\beta_{21}$ ) estimates from Table 24 and Table 25 for separate and joint models for simulation set 1, investigating varying association parameters. The dashed line identifies the “true” value of  $\beta_{21}$  that the data was simulated under.

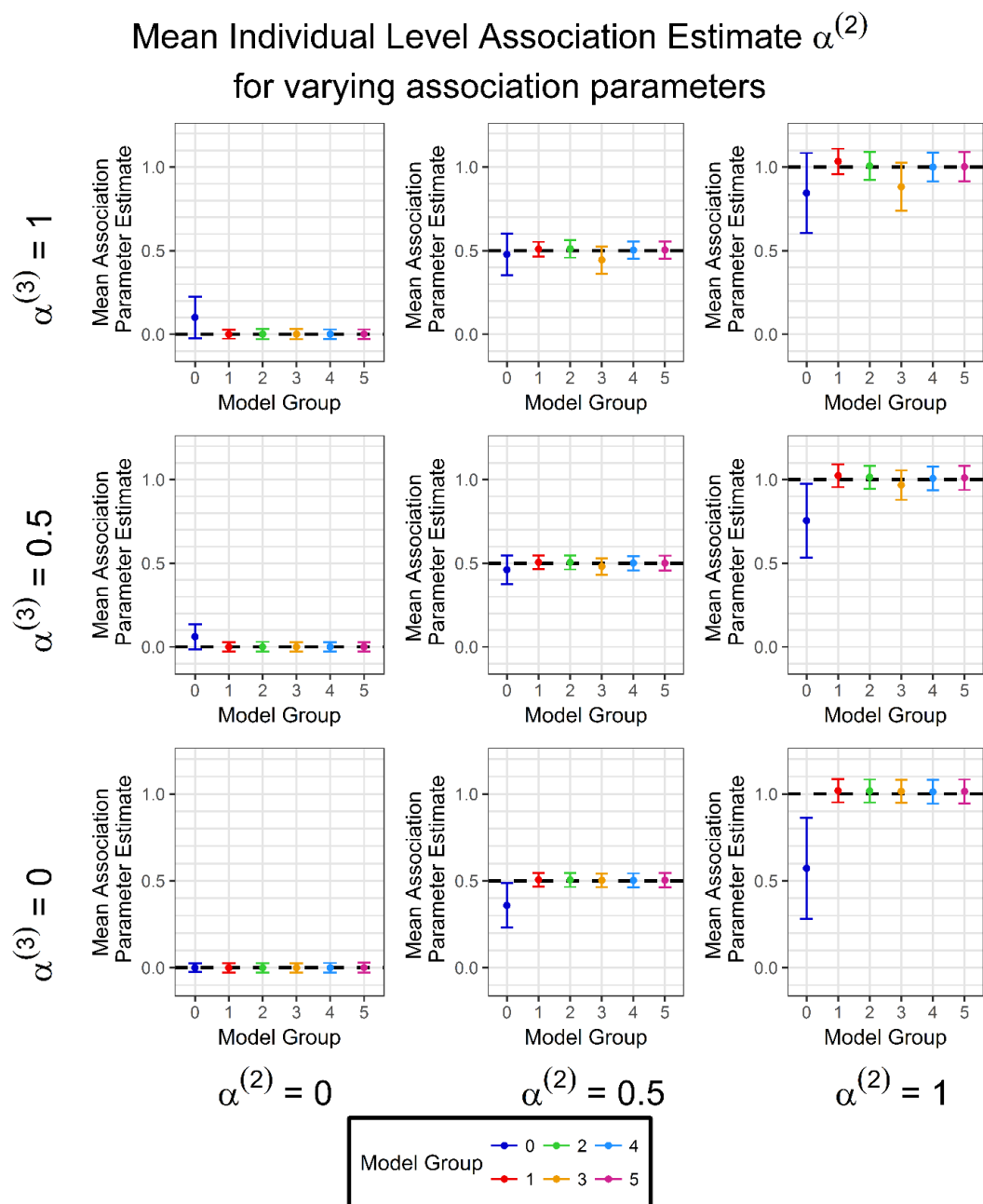


Figure 27: Graphical representation of the mean individual level association parameter ( $\alpha^{(2)}$ ) estimates from Table 24 and Table 25 for separate and joint models for simulation set 1, investigating varying association parameters. The dashed line identifies the “true” value of  $\alpha^{(2)}$  that the data was simulated under.

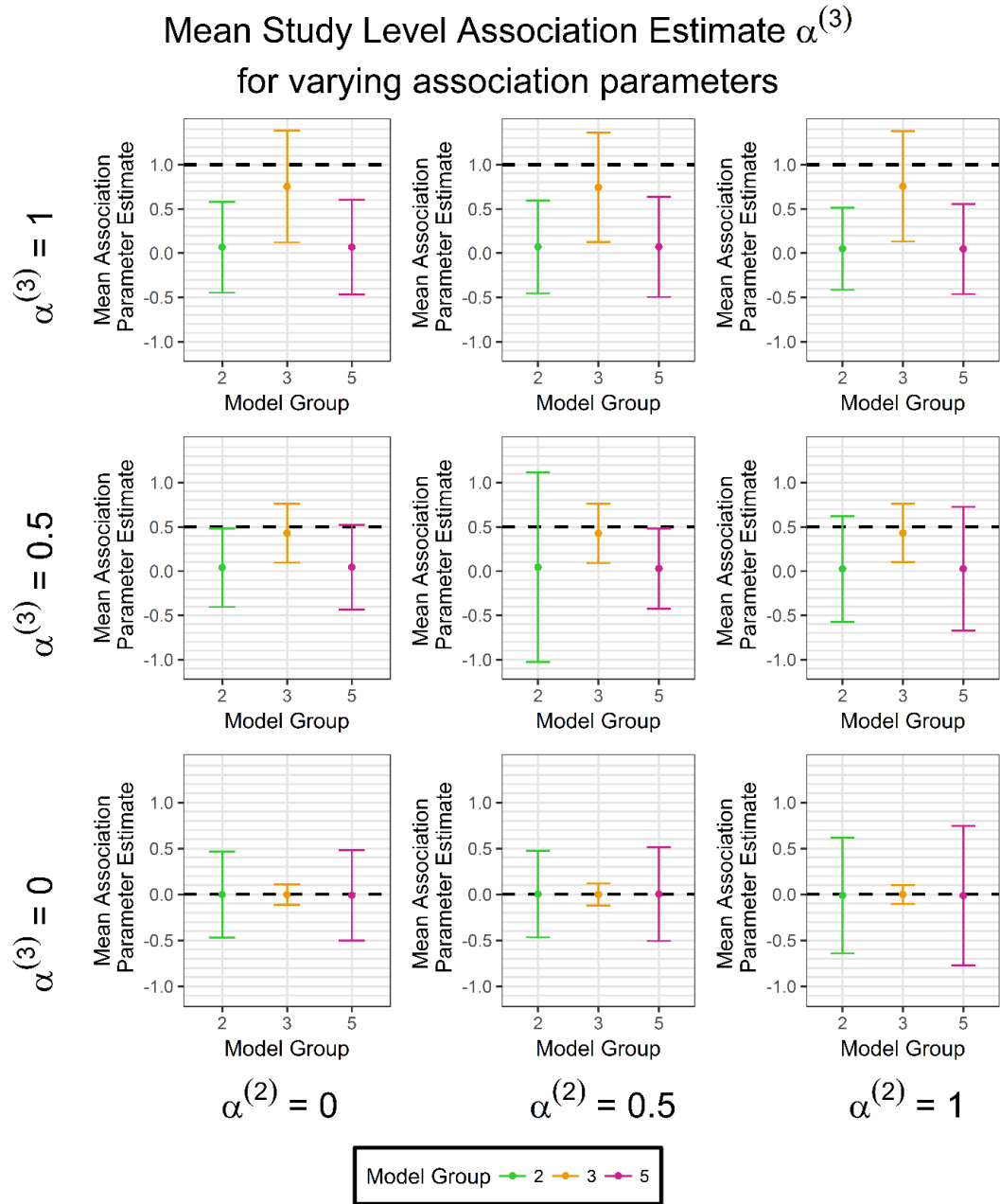


Figure 28: Graphical representation of the mean study level association parameter ( $\alpha^{(3)}$ ) estimates from Table 24 and Table 25 for separate and joint models for simulation set 1, investigating varying association parameters. The dashed line identifies the “true” value of  $\alpha^{(3)}$  that the data was simulated under.

### 7.2.3.2 *Simulation Set 2: Varying numbers of studies included in the meta-analysis*

For this second set of simulations (Table 26 and Figure 29-Figure 32), the lowest proportion of successfully completed model fits was 99.9% (one failed fit for each of model groups 3 and 5).

Across all scenarios and model groups, for both the separate and joint analyses, the mean estimate of the longitudinal treatment effect  $\beta_{12}$  was close to the “true” value of 2 (Table 26, Figure 29). The SE were comparable across model groups and scenarios between the separate and joint analyses, with SE of the simulation estimates decreasing as the number of included studies increased (Figure 29). Coverage was poor for both the separate and joint analyses across scenarios for the naïve model group (group 0). For the remaining model groups, again the separate models showed good coverage (between 86 and 94%) across investigated scenarios. However, joint model results from model groups involving study level random effects again showed poor coverage (between 9 and 20%), whilst those from model groups not involving study level random effects showed coverage between 87 and 94%. Again the reason for this can be seen through examination of Figure 52, Appendix 4; again whilst the point estimates for  $\beta_{12}$  from model groups 2, 3 and 5 are clustered around the true parameter value, the confidence intervals are narrow.

For all scenarios for all model groups investigated, the separate analyses underestimated the true value of the time-to-event treatment effect  $\beta_{21}$ , whereas the mean estimate from the joint models was closer to the true value of 3 (Table 26, Figure 30). The SE of the estimates decreased as the number of included studies increased for both the separate and joint analyses (Table 26, Figure 30), although the SE for the separate analyses was often smaller than that of the joint analyses. Coverage for the estimates from the separate analyses were low for all model groups and scenarios (6% or less). Coverage for the joint model analyses was higher across all model groups, although still much lower than the ideal level of 95% (Table 26). Examination of Figure 53, Appendix 4, indicates that despite the reduced coverage due to narrow confidence intervals, for model groups involving study level random effects, the estimates from joint models were clustered about the true value.

The individual level association parameter  $\alpha^{(2)}$  was estimated in every fitted joint model due to the presence of individual level random effects in each model. For any model group that accounted for between study heterogeneity, the estimates of  $\alpha^{(2)}$  were close to the true value of 0.5 with low SE (Table 26, Figure 31). Coverage was good (between 86 and



95%) for model groups 1, 2, 4 and 5. However, model group 3 (which accounted for between study heterogeneity solely through study level random effects) had coverages that decreased as the number of included studies increased. In addition the naïve model group (group 0) had coverage of less than 50% for all scenarios investigated. Examination of the simulation specific confidence intervals in Figure 54, Appendix 4, shows that for model groups 0 and 3, as the number of included studies increased, the area covered by the confidence intervals of the estimates narrowed, leading to poorer coverage.

The study level association parameter  $\alpha^{(3)}$  was poorly estimated in all model groups it appeared in (2, 3 and 5), across all scenarios. The estimates in model groups 2 and 5 (which included a study level random treatment effect) were close to zero, with SE and coverage decreasing as the number of included studies increased (Table 26, Figure 32). The estimates for model group 3 were closer to the true value of 0.5, however the coverage was still poor (between 29 and 43% for all scenarios). Examination Figure 55, Appendix 4, shows confidence intervals decreasing in width for all model groups as the number of studies increases, with wider confidence intervals for model groups 2 and 5. However, only the estimates for model group 3 are clustered close to the true value.

	Number of included studies	Number of successful joint fits	Longitudinal Treatment Effect ( $\beta_{12} = 2$ )		Time-to-event treatment effect ( $\beta_{21} = 3$ )		Association Parameters	
			Separate model	Joint Model	Separate Model	Joint Model	Joint Model ( $\alpha^{(2)} = 0.5$ )	Joint Model ( $\alpha^{(3)} = 0.5$ )
Group 0	5	1000	2.01 (0.55) [18.5]	2.01 (0.55) [17.0]	1.57 (0.18) [0.0]	2.74 (0.29) [43.5]	0.461 (0.04) [49.7]	NA
	10	1000	2.00 (0.38) [19.5]	2.00 (0.38) [18.8]	1.54 (0.12) [0.0]	2.69 (0.22) [28.4]	0.457 (0.03) [31.3]	NA
	15	1000	2.01 (0.31) [22.6]	2.01 (0.31) [20.1]	1.54 (0.10) [0.0]	2.66 (0.18) [14.6]	0.452 (0.03) [17.5]	NA
Group 1	5	1000	2.01 (0.55) [86.2]	2.01 (0.55) [87.5]	1.69 (0.47) [6.0]	3.04 (0.26) [85.0]	0.506 (0.02) [92.8]	NA
	10	1000	2.00 (0.38) [92.2]	2.00 (0.38) [92.6]	1.69 (0.50) [3.2]	3.04 (0.18) [89.6]	0.507 (0.01) [88.0]	NA
	15	1000	2.01 (0.31) [93.9]	2.01 (0.31) [93.5]	1.69 (0.47) [1.9]	3.04 (0.15) [91.7]	0.506 (0.01) [86.0]	NA
Group 2	5	1000	2.01 (0.55) [87.3]	2.01 (0.55) [11.0]	1.76 (0.21) [0.0]	3.10 (0.27) [62.5]	0.505 (0.02) [93.7]	0.045 (0.55) [32.3]
	10	1000	2.00 (0.38) [92.2]	2.00 (0.38) [9.8]	1.76 (0.15) [0.0]	3.11 (0.19) [59.1]	0.506 (0.02) [90.4]	0.035 (0.12) [5.3]
	15	1000	2.01 (0.31) [93.3]	2.01 (0.31) [12.2]	1.76 (0.12) [0.0]	3.10 (0.15) [56.4]	0.505 (0.01) [90.6]	0.029 (0.09) [0.5]
Group 3	5	999	2.01 (0.55) [87.4]	2.01 (0.55) [11.4]	1.57 (0.18) [0.0]	2.88 (0.27) [62.9]	0.481 (0.03) [83.2]	0.427 (0.17) [42.8]
	10	1000	2.00 (0.38) [92.2]	2.00 (0.38) [12.2]	1.54 (0.12) [0.0]	2.83 (0.19) [52.5]	0.474 (0.02) [60.9]	0.426 (0.10) [38.2]
	15	1000	2.01 (0.31) [93.3]	2.01 (0.31) [13.9]	1.54 (0.10) [0.0]	2.80 (0.16) [39.3]	0.471 (0.02) [39.5]	0.416 (0.08) [29.5]
Group 4	5	1000	2.01 (0.54) [86.5]	2.01 (0.54) [87.5]	1.73 (0.20) [0.0]	3.04 (0.26) [69.7]	0.501 (0.02) [95.2]	NA
	10	1000	2.00 (0.38) [92.2]	2.00 (0.38) [92.3]	1.73 (0.14) [0.0]	3.04 (0.18) [69.7]	0.502 (0.02) [94.0]	NA
	15	1000	2.01 (0.31) [93.9]	2.01 (0.31) [93.5]	1.73 (0.11) [0.0]	3.03 (0.15) [70.6]	0.501 (0.01) [94.8]	NA
Group 5	5	999	2.04 (0.58) [84.8]	2.04 (0.58) [10.6]	1.74 (0.20) [0.0]	3.04 (0.26) [70.2]	0.501 (0.02) [94.8]	0.030 (0.23) [41.3]
	10	1000	2.00 (0.38) [92.2]	2.00 (0.38) [10.7]	1.73 (0.14) [0.0]	3.04 (0.18) [69.2]	0.503 (0.02) [93.9]	0.038 (0.12) [9.2]
	15	1000	2.01 (0.31) [93.3]	2.01 (0.31) [12.3]	1.73 (0.11) [0.0]	3.03 (0.15) [69.5]	0.501 (0.01) [93.6]	0.032 (0.10) [1.4]

Table 26: Results for simulations for longitudinal treatment effect parameter  $\beta_{12}$ , time-to-event treatment parameter  $\beta_{21}$ , and association parameters ( $\alpha^2$  and if estimated  $\alpha^{(3)}$ ) for varying numbers of included studies. Results stated as mean parameter estimate (SE between simulation estimates) [coverage]. Model group 1 reports overall longitudinal and time-to-event treatment effect estimates, and model group 4 reports overall longitudinal treatment effect estimates produced by combining study specific treatment effect estimates (see section 7.2.2). NA indicates models that did not estimate  $\alpha^{(3)}$ . "True" parameter values are stated in the column headers.

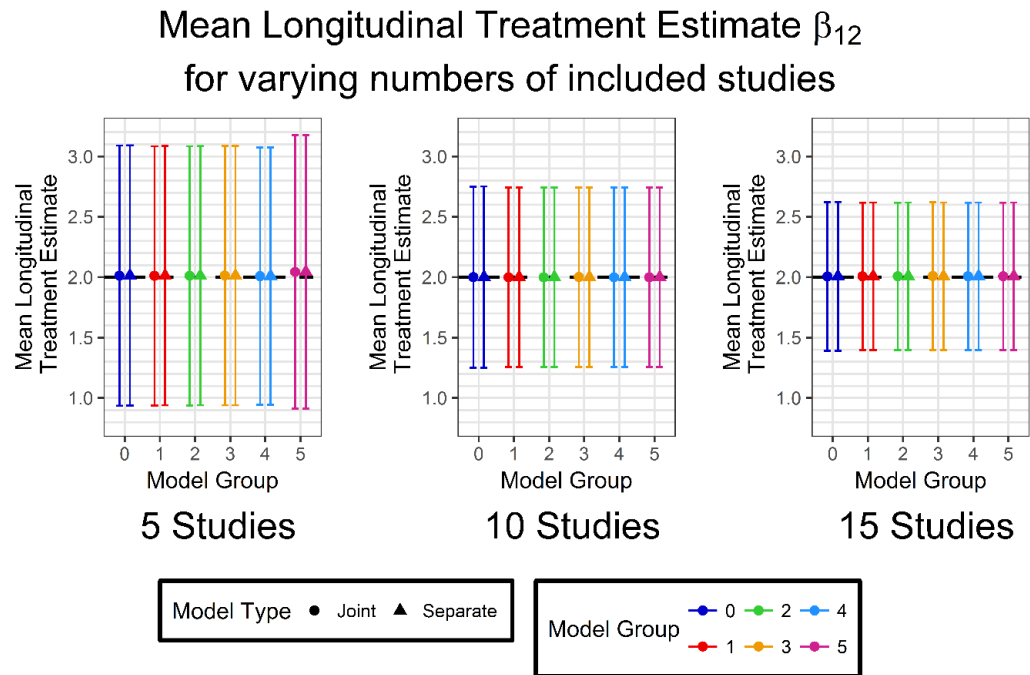


Figure 29: Graphical representation of longitudinal treatment effect ( $\beta_{12}$ ) estimates shown in Table 26 for simulation set 2: investigation of varying numbers of included studies. The dashed line indicates the “true” value of  $\beta_{12}$  that the data was simulated under

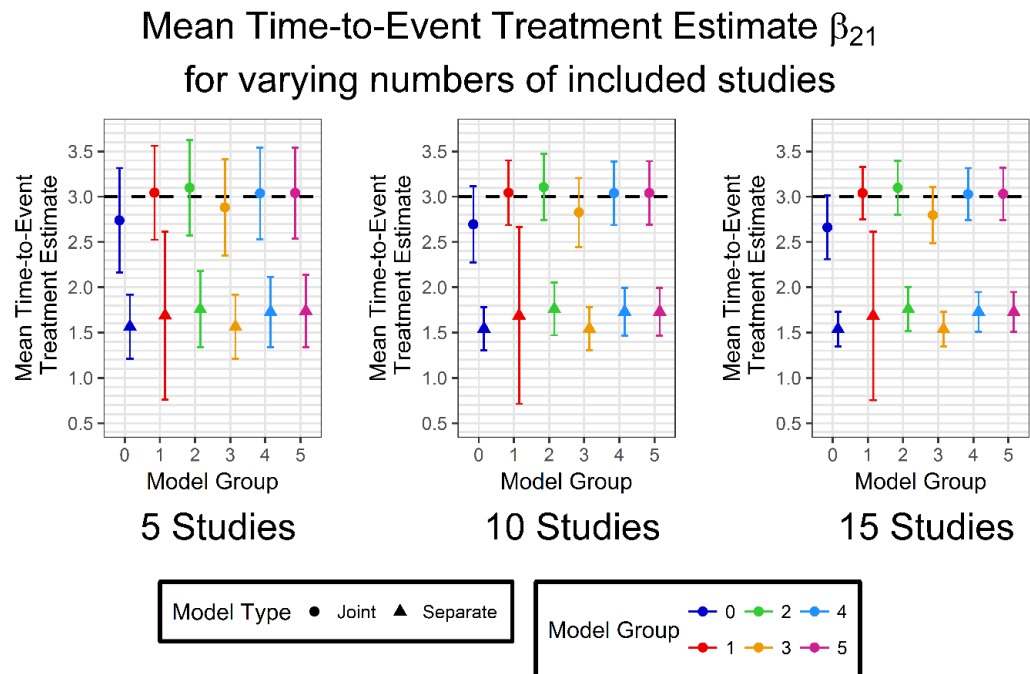


Figure 30: Graphical representation of time-to-event treatment effect ( $\beta_{21}$ ) estimates shown in Table 26 for simulation set 2: investigation of varying numbers of included studies. The dashed line indicates the “true” value of  $\beta_{21}$  that the data was simulated under

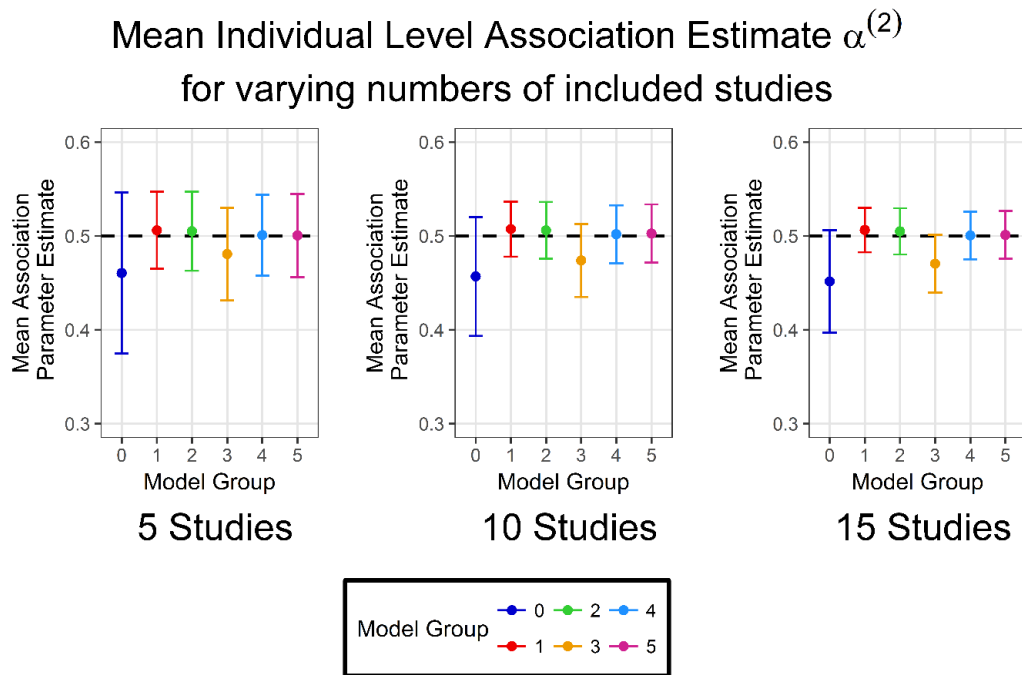


Figure 31: Graphical representation of individual level association parameter ( $\alpha^{(2)}$ ) estimates shown in Table 26 for simulation set 2: investigation of varying numbers of included studies. The dashed line indicates the “true” value of  $\alpha^{(2)}$  that the data was simulated under

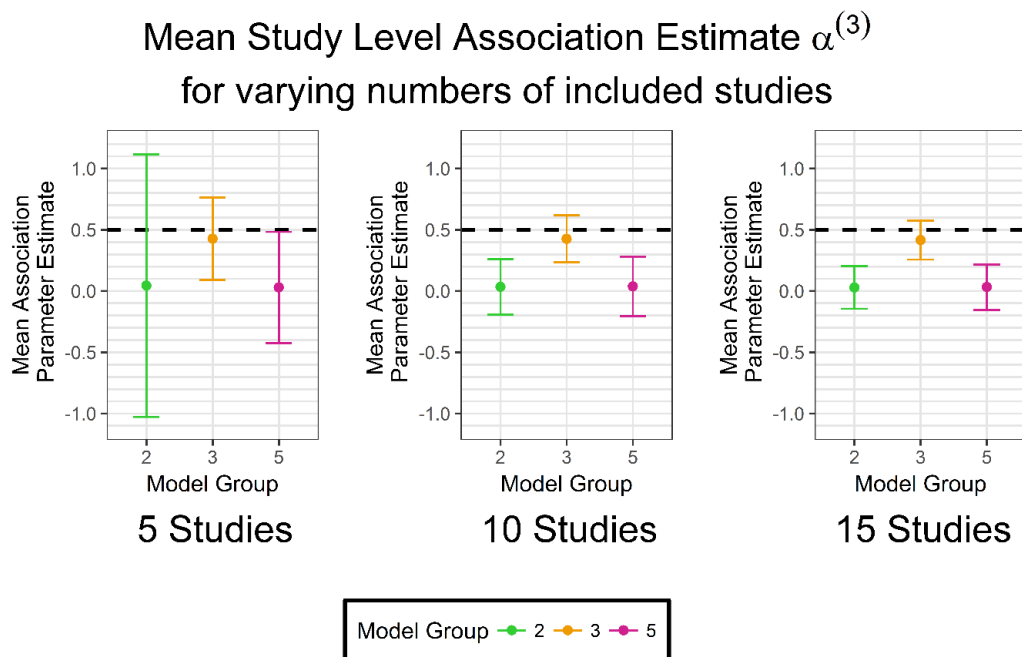


Figure 32: Graphical representation of study level association parameter ( $\alpha^{(3)}$ ) estimates shown in Table 26 for simulation set 2: investigation of varying numbers of included studies. The dashed line indicates the “true” value of  $\alpha^{(3)}$  that the data was simulated under

### 7.2.3.3 *Simulation Set 3: Varying levels of between study heterogeneity*

For this third set of simulations (Table 27 and Figure 33-Figure 36), the proportions of successful model fits were low for model groups involving study level random effects (2, 3 and 5) where there was no between study heterogeneity ( $A = A_1 = 0$ ). Otherwise, the proportion of successful model fits was at least 99.8%.

The estimates from the separate and joint analyses for all model groups for all scenarios were close to the “true” value of 2 for the longitudinal treatment effect  $\beta_{12}$  (Table 27, Figure 33). SE was similar in all cases between the separate and joint analyses, however the coverage differed noticeably. For any model group that accounted for between study heterogeneity, coverage was above 85% for the separate analyses. However, for the joint models, whilst model groups 1 and 4 showed coverage above 88% across all scenarios, any model group involving study level random effects, coverage dropped sharply for non-zero between study heterogeneity. Examining Figure 56, Appendix 4, the point estimates from the joint analyses can be seen to be clustered around the true value, however the confidence intervals are narrow leading to low coverage.

The mean estimates from the joint analyses of the time-to-event treatment effect  $\beta_{21}$  are close to the “true” value of 3, however fall below the “true” value for the separate analyses (Table 27, Figure 34). The SE are higher for the joint analyses, with SE increasing as between study heterogeneity increases. Coverage for the separate analyses is poor (below 14% for all model groups for all scenarios, with most coverages at 0%). For the joint analyses, coverage for model group 1 is constant at around 86%. For model group 3, the coverage is lower (below 63% for all scenarios). For the remaining model groups for the joint analyses, the coverage is high (above 93%) for cases with no between study heterogeneity, but coverage decreases as the level of between study heterogeneity increases. From Figure 57, Appendix 4, the width of the band in which the point estimates and their confidence intervals falls in widens (most noticeably for model group 1) as between study heterogeneity increases. Given the increase in the SE of the point estimates, the coverage may be dropping due to more variable point estimates, rather than changes in confidence interval width.

The mean estimates for the individual level association parameter  $\alpha^{(2)}$  are close to the “true” value of 0.5, with low SE, across all model groups and scenarios (Table 27, Figure 35). However, the coverage for model group 0 decreases as the between study heterogeneity increases. Otherwise, for the remaining model groups (except group 3) coverage remains

high (above 90%) regardless of the level of between study heterogeneity (although in several cases it decreases slightly as between study heterogeneity increases). For model group 3, the coverage is variable, but does not appear to follow a trend as between study heterogeneity increases. Examination of Figure 58, Appendix 4, shows the estimates from the model groups that account for between study heterogeneity clustered around the true value of the estimate, regardless of coverage level.

The study level association parameter  $\alpha^{(3)}$  is estimated poorly in model groups 2 and 5, with estimates negative or close to zero (Table 27, Figure 36). The estimates from model group 3 are closer to the “true” value, but are still underestimated. The SE of the estimates is large when there is no between study heterogeneity, and decrease as the level of between study heterogeneity increases. The coverage decreases as the between study heterogeneity increases. Figure 59, Appendix 4, clearly shows the poor parameter estimation for no between study heterogeneity, and shows the point estimates for model group 3 clustered about the true value, however those for model groups 2 and 5 clustered slightly below the true value.

Study level/ covariance matrix	Number of successful joint fits	Longitudinal Treatment Effect ( $\beta_{12} = 2$ )		Time-to-event treatment effect ( $\beta_{21} = 3$ )		Association Parameters	
		Separate model	Joint Model	Separate Model	Joint Model	Joint Model ( $\alpha^{(2)} = 0.5$ )	Joint Model ( $\alpha^{(3)} = 0.5$ )
Group 0	$A = A_1$	2.00 (0.04) [95.2]	2.00 (0.04) [94.8]	1.63 (0.11) [0.0]	3.01 (0.14) [93.7]	0.502 (0.02) [93.4]	NA
	$A = A_2$	2.01 (0.55) [18.5]	2.01 (0.55) [17.0]	1.57 (0.18) [0.0]	2.74 (0.29) [43.5]	0.461 (0.04) [49.7]	NA
	$A = A_3$	1.99 (0.79) [17.3]	1.99 (0.79) [16.1]	1.52 (0.22) [0.0]	2.57 (0.40) [28.0]	0.435 (0.06) [32.8]	NA
Group 1	$A = A_1$	2.00 (0.04) [97.2]	2.00 (0.04) [96.6]	1.65 (0.18) [0.0]	3.03 (0.14) [87.5]	0.506 (0.02) [91.7]	NA
	$A = A_2$	2.01 (0.55) [86.2]	2.01 (0.55) [87.5]	1.69 (0.47) [6.0]	3.04 (0.26) [85.0]	0.506 (0.02) [92.8]	NA
	$A = A_3$	2.00 (0.79) [88.4]	1.99 (0.79) [88.4]	1.69 (0.63) [13.3]	3.05 (0.35) [85.9]	0.508 (0.02) [90.5]	NA
Group 2	$A = A_1$	2.01 (0.04) [100.0]	2.00 (0.04) [97.4]	1.64 (0.12) [0.0]	3.05 (0.15) [93.4]	0.512 (0.02) [93.4]	-0.377 (8.86) [100.0]
	$A = A_2$	2.01 (0.55) [87.3]	2.01 (0.55) [11.0]	1.76 (0.21) [0.0]	3.10 (0.27) [62.5]	0.505 (0.02) [93.7]	0.045 (0.55) [32.3]
	$A = A_3$	2.00 (0.79) [88.5]	1.99 (0.79) [7.9]	1.85 (0.29) [0.0]	3.16 (0.35) [50.7]	0.508 (0.02) [90.7]	0.027 (0.16) [15.2]
Group 3	$A = A_1$	2.00 (0.03) [97.0]	2.00 (0.03) [45.3]	1.63 (0.11) [0.0]	3.02 (0.14) [44.3]	0.505 (0.02) [44.3]	0.380 (2.64) [47.3]
	$A = A_2$	2.01 (0.55) [87.4]	2.01 (0.55) [11.4]	1.57 (0.18) [0.0]	2.88 (0.27) [62.9]	0.481 (0.03) [83.2]	0.427 (0.17) [42.8]
	$A = A_3$	2.00 (0.79) [88.5]	2.00 (0.79) [8.6]	1.52 (0.22) [0.0]	2.78 (0.36) [47.3]	0.464 (0.03) [65.1]	0.403 (0.16) [31.8]
Group 4	$A = A_1$	2.00 (0.04) [97.2]	2.00 (0.04) [96.7]	1.67 (0.11) [0.0]	3.01 (0.14) [94.7]	0.502 (0.02) [95.3]	NA
	$A = A_2$	2.01 (0.54) [86.5]	2.01 (0.54) [87.5]	1.73 (0.20) [0.0]	3.04 (0.26) [69.7]	0.501 (0.02) [95.2]	NA
	$A = A_3$	2.00 (0.79) [88.4]	1.99 (0.79) [88.6]	1.78 (0.26) [0.0]	3.06 (0.33) [56.9]	0.502 (0.02) [93.6]	NA
Group 5	$A = A_1$	2.00 (0.04) [100.0]	2.00 (0.04) [100.0]	1.66 (0.12) [0.0]	3.04 (0.17) [94.3]	0.509 (0.02) [96.2]	-1.343 (6.03) [100.0]
	$A = A_2$	2.04 (0.58) [84.8]	2.04 (0.58) [10.6]	1.74 (0.20) [0.0]	3.04 (0.26) [70.2]	0.501 (0.02) [94.8]	0.030 (0.23) [41.3]
	$A = A_3$	2.00 (0.79) [88.5]	1.99 (0.79) [7.8]	1.78 (0.26) [0.0]	3.07 (0.33) [56.4]	0.503 (0.02) [93.5]	0.030 (0.17) [19.8]

Table 27: Results for simulations for longitudinal treatment effect parameter  $\beta_{12}$ , time-to-event treatment parameter  $\beta_{21}$ , and association parameters ( $\alpha^{(2)}$  and if estimated  $\alpha^{(3)}$ ) for varying levels of heterogeneity between studies. Results stated as mean parameter estimate (SE between simulation estimates) [coverage]. Model group 1 reports overall longitudinal and time-to-event treatment effect estimates, and model group 4 reports overall longitudinal treatment effect estimates produced by combining study specific treatment effect estimates (see Section 7.2.2). Definitions of  $A_1$ ,  $A_2$  and  $A_3$  are given in Table 22. NA identifies model groups which did not estimate the  $\alpha^{(3)}$  parameter. "True" parameter values are stated in the column headers

### Mean Longitudinal Treatment Estimate $\beta_{12}$ for varying numbers of included studies

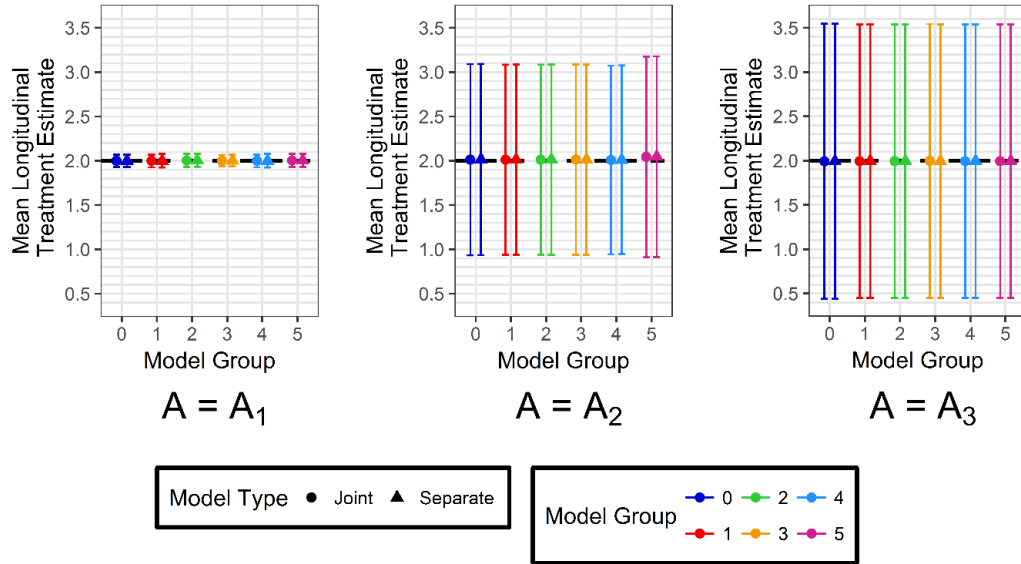


Figure 33: Graphical representation of longitudinal treatment effect ( $\beta_{12}$ ) estimates shown in Table 27 for simulation set 3: investigation of varying between study heterogeneity. The dashed line identifies the "true" value of  $\beta_{12}$  that the data was simulated under

### Mean Time-to-Event Treatment Estimate $\beta_{21}$ for varying levels of between study heterogeneity

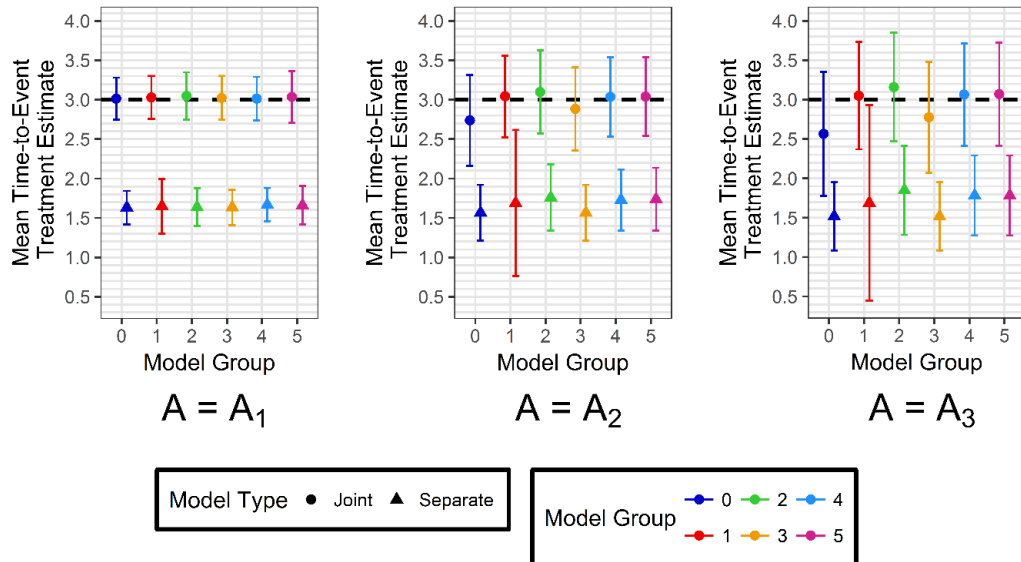


Figure 34: Graphical representation of time-to-event treatment effect ( $\beta_{21}$ ) estimates shown in Table 27 for simulation set 3: investigation of varying between study heterogeneity. The dashed line identifies the "true" value of  $\beta_{21}$  that the data was simulated under



### Mean Individual Level Association Estimate $\alpha^{(2)}$ for varying levels of between study heterogeneity

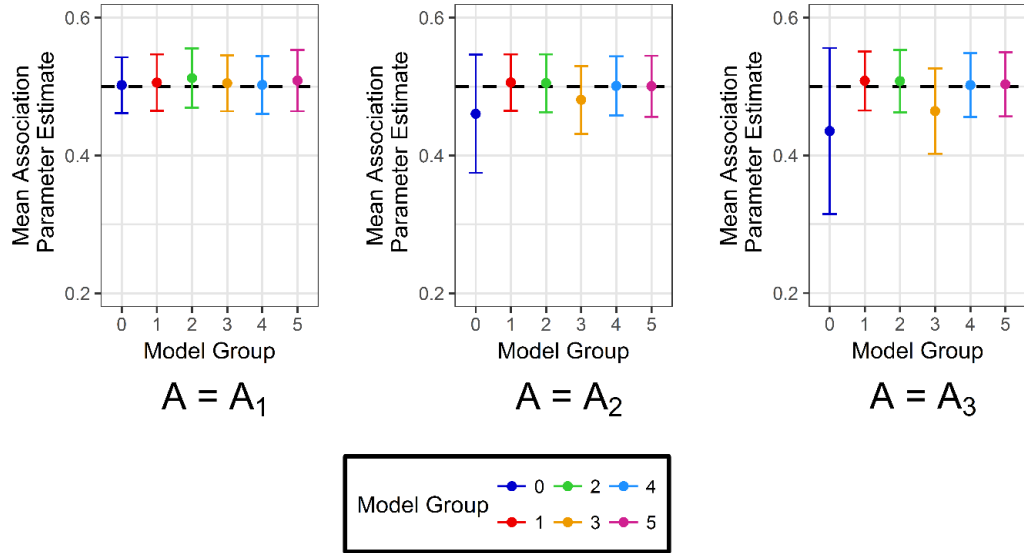


Figure 35: Graphical representation of individual level association parameter ( $\alpha^{(2)}$ ) estimates shown in Table 27 for simulation set 3: investigation of varying between study heterogeneity. The dashed line identifies the "true" value of  $\alpha^{(2)}$  that the data was simulated under

### Mean Study Level Association Estimate $\alpha^{(3)}$ for varying levels of between study heterogeneity

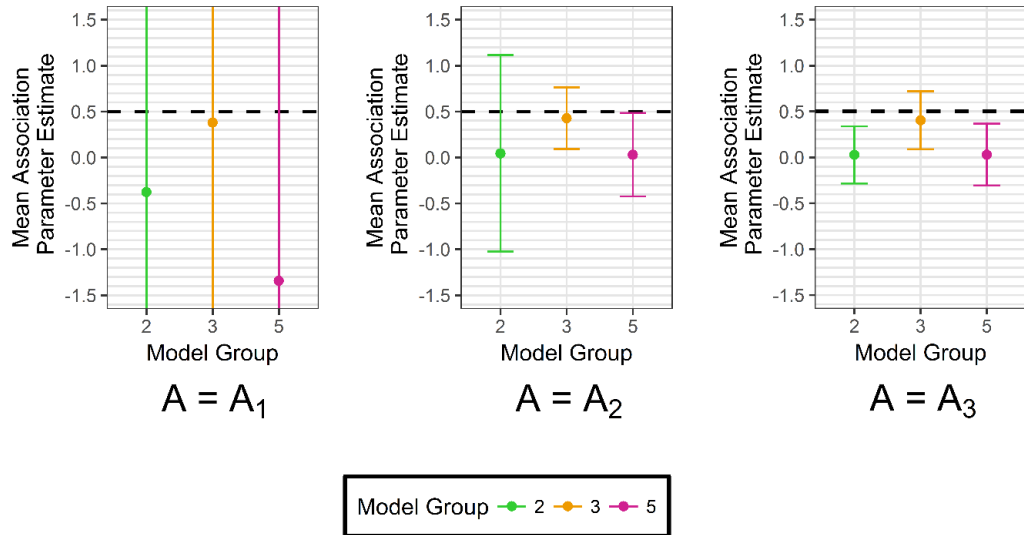


Figure 36: Graphical representation of study level association parameter ( $\alpha^{(3)}$ ) estimates shown in Table 27 for simulation set 3: investigation of varying between study heterogeneity. The dashed line identifies the "true" value of  $\alpha^{(3)}$  that the data was simulated under

## 7.3 Discussion of simulation studies

During this chapter, simulation studies investigating the methods presented in Chapter 3 have been presented. A range of scenarios, for both one and two-stage methods, have been explored.

### 7.3.1 Discussion of the Two-Stage Simulation Investigation

The real data two-stage MA reported in Section 6.2 did not show sizeable differences between MA that employed separate methods and MA that employed joint methods in the first stage of the MA. However, the simulation studies involving larger magnitude associations (Section 7.1) showed more apparent differences between pooled results from separate time-to-event analyses, and joint analyses.

The fixed and random MA approaches generally gave similar results for the homogenous simulated datasets, but the results differed more for the heterogeneous treatment simulated datasets. This suggests that researchers conducting a two-stage MA using joint longitudinal and time-to-event data should, as with other MA, ensure that random effects methods are used if it is known or suspected that heterogeneity exists between the trials included in the meta-analysis. This can be achieved through examination of forest plots, and exploring estimates of the  $\tau^2$  parameter and  $I^2$  statistic.

Throughout the two-stage simulations, the coverage for the joint modelling approach appeared consistently high across different scenarios and association levels. Examination of the estimates from the model fits in the simulation study highlighted that the joint approach consistently gave pooled estimates close to the true values for association parameters or treatment effects from either sub-model.

The simulation study also identified that the separate longitudinal approach appeared adequate across scenarios investigated, for any association level examined. However, whilst the separate time-to-event approach displayed little bias for pooled treatment effect where association between the time-to-event and longitudinal outcomes was insignificant, where a positive association was present, the treatment effect was underestimated, with bias increasing as the strength of association increases. This underestimation was observed only in the separate time-to-event analyses, not in the joint analyses, and was not restricted to the two-stage approach. This is discussed further below.

### 7.3.2 Discussion of the One-Stage Simulation Investigation

As with the two-stage simulation study, similarities existed between the estimates of longitudinal treatment effect between the separate and joint one-stage MA models across the scenarios investigated (Section 7.2). However, where non-zero positive association existed at the individual level, separate time-to-event one-stage MA models underestimated the true treatment effect by an increasing amount as individual level association increased. In comparison, the one-stage joint MA models that accounted for between study heterogeneity estimated the time-to-event treatment effect satisfactorily across all association levels tested. The naïve one-stage MA models (joint or separate) that ignored between study heterogeneity behaved poorly across the scenarios examined where association was present between the longitudinal and time-to-event outcomes.

Joint models that shared study level random effects between sub-models through the association structure (model groups 2, 3 and 5) displayed issues. These three model groups produced longitudinal treatment effect estimates that clustered about the “true” parameter value, but displayed poor coverage. However, separate longitudinal models with the same specification displayed similar treatment effect estimates, with improved coverage. Estimation of the study level association parameter ( $\alpha^{(3)}$ ) was poor for model groups 2 and 5, and slightly better for model group 3 (which accounted for between study heterogeneity solely through study level random effects). However, model group 3 displayed issues in the estimation of other parameters, e.g. the time-to-event treatment effect  $\beta_{21}$  and the individual level association parameter  $\alpha^{(2)}$ . The behaviour of joint model groups involving study level random effects was not improved for differing levels of between study heterogeneity, or different numbers of included studies.

Consequently, I recommend, for scenarios similar to those investigated here, when one-stage joint MA models are being employed, not to employ methods that share study level random effects between sub-models. However, the separate longitudinal models that involved study level random effects appeared to avoid some of the problems encountered with the joint approach. It may be that study level random effects could be employed in the longitudinal sub-model, with between study heterogeneity accounted for in the time-to-event sub-model through stratification of the baseline hazard, fixed interaction terms with the study membership variable, or by a frailty term, without any sharing of study level random effects between sub-models. This avenue of future research is discussed in Section 8.3.2.2.

From the currently investigated one-stage joint MA model specifications, unless the number of included studies renders such an analysis unwieldy, the simulation studies indicate model group 1 (which includes fixed interaction terms with the study membership in both the longitudinal and time-to-event sub-models) to have the best overall coverage, followed by model group 4 (which involves fixed interaction terms with the study membership variable in the longitudinal sub-model, and baseline hazard stratified by study with study membership as a fixed effect in the time-to-event sub-model). Time-to-event parameters were better estimated in model group 1 than 4 where there was non-zero study level association, however model group 4 had marginally faster fitting times (as it involved a stratified baseline hazard, meaning that model fitting involved matrices of smaller dimensions than unstratified approaches).

### 7.3.3 Comparison of one and two-stage simulations

As noted in Sections 3.4 and 6.4.3, one-stage methods are beneficial compared to two-stage methods in that their likelihood specification is more accurate [189-191], however they are noted to be more computationally intensive to conduct [63, 189, 191]. An example of this in the simulation studies was in the estimation of the baseline hazard. In the study specific model fits involved in the two-stage analyses, estimation of the baseline hazard involved only the event times observed in that study's population. One-stage fits that did not stratify the baseline hazard by study used event times across all studies, resulting calculations using larger matrices and vectors, which were more computationally intensive.

It is important to note the differences in the simulation of multi-study joint datasets between the one and two-stage simulation studies, which could cause differences in the results from the two approaches. During the two-stage simulations, homogenous datasets were created by simulating data for each study under the same model parameters (with minor variability between the dataset for each study stemming from the use of probability distributions to generate error terms and individual level random effects). The between study heterogeneity in the two-stage simulation heterogeneous datasets was created by generating model parameters from normal distributions, whose mean value equalled the "true" parameter value.

Conversely, during the one-stage simulations, between study heterogeneity was controlled using the distribution of the study level random effects; larger values in their covariance matrix causes greater between study variability in both the longitudinal and time-to-event

outcomes. However, scenarios that varied the study level association parameter ( $\alpha^{(3)}$ ) also inherently varied the level of between study heterogeneity. Larger magnitudes of  $\alpha^{(3)}$  increased the effect of the study level random effects on risk of the event, magnifying the between study heterogeneity in the time-to-event outcome.

In addition, the time points at which longitudinal measurements were recorded differed between the datasets simulated for the one and the two-stage investigations. This difference stemmed from early issues with the one-stage modelling code: initially it was thought that convergence issues may be attributable to insufficient data, leading to inflation of the number of longitudinal time-points in datasets simulated for use during development of the *jointmeta1* function, however introduction of pseudo-adaptive quadrature methods resolved the convergence issues (see Section 5.1.2.2.1).

Despite the mentioned differences in multi-study joint data simulation, the one and two-stage simulations agree on several points. Firstly, the longitudinal effect estimates produced by meta-analytic joint methods and the meta-analytic separate longitudinal methods are similar (although one-stage joint models that involved study level random effects reported poor coverage). Secondly, where non-zero individual level association is present between the longitudinal and time-to-event outcomes, separate time-to-event meta-analytic models produce biased estimates of parameters, whereas the estimates produced by joint meta-analytic models are closer to the “true” parameter values.

This behaviour may be linked to the known case where omission of covariates from Cox models leads to bias in estimated effect parameters [241-243]. Compared to the joint time-to-event sub-model (equation (70) and Table 23), the separate time-to-event models do not include the  $W_{2ki}(t)$  term. Where association is present (i.e. when  $\alpha_k^{(2)} \neq 0$  for the two-stage, or when  $\alpha^{(2)}$  or  $\alpha^{(3)} \neq 0$  for the one-stage) the joint approach models risk of an event associated with the longitudinal outcome via  $W_{2ki}(t)$ . This term (which has an effect on the event risk) is not included in the separate time-to-event model, potentially explaining the observed biased treatment effect estimates. This disparity in parameter point estimates was not observed between the separate and joint longitudinal analyses as the model specifications for the longitudinal trajectory are identical in both models.

As noted, at the start of this research, there was a gap in the literature concerning joint modelling of multi-study data. However, Guo and Carlin [158] analysed a single-study dataset using separate and joint models. The joint model association parameter was significant and negative. The differences observed between the estimated survival times

produced by the separate and the joint models was attributed to the fact that the joint models accounted for correlation between the longitudinal and time-to-event outcomes, whereas the separate time-to-event model did not.

In the simulations reported in Chapter 7, the parameter estimates produced by the standalone time-to-event model differ to those produced by the joint model (potential reasons for this have been discussed in previous paragraphs). However the results from standalone longitudinal models and the joint model were similar. There may be cases where results from standalone longitudinal models and joint models are significantly different. An example of this could be the case of longitudinal data where patients who do not respond as well to treatment as others in the population are more likely to drop out. In this case, if dropout is not accounted for in the model (i.e. a standalone longitudinal model is fitted) estimates of treatment effectiveness on the longitudinal outcome are based only on the healthier patients, and so may be biased. A joint analysis that accounts for outcome related dropout may produce different effect estimates. Future research specifically into longitudinal data complicated by outcome related dropout would be beneficial.

Whilst no multi-study joint data simulation studies were identified, many single study investigations employed simulation approaches to test joint modelling methodology. For example, Sweeting and Thompson [244] performed simulation studies to compare three methods to model patients with abdominal aortic aneurysms: time-to-event models with time-varying covariates, a two-stage approach (where a longitudinal model was fitted, and the longitudinal fitted values then included in a time-to-event model), and joint modelling methods. Although the data was multi-centre, this does not appear to have been accounted for in their models, and so the simulations reflect a single study case. Their simulation study concluded that the joint modeling approach was preferred, regardless of the higher computational burden.

In this simulation investigation a similar result was observed for the multi-study case for both one and two-stage approaches – where an individual level association between the longitudinal and time-to-event outcomes existed, the pooled estimates from the separate time-to-event model underestimated the true simulated time-to-event treatment effect compared to the pooled estimates from the joint model. However, where this association was insignificant, the separate and joint analyses produced similar results. Therefore, as with single study cases, in one or two-stage meta-analytic case there is evidence of benefit of joint methods over separate methods for estimation of time-to-event coefficients where

an association between the longitudinal and time-to-event outcomes is known or suspected. As discussed in Section 3.2.1, plotting the longitudinal outcome against time ( $t_{kij}$ ) minus survival time ( $T_{ski}$ ) for each individual, panelled by event type (censored or experienced the event), highlights changes immediately before the individual experiences the event in question, or is censored. This is beneficial, as differences in the shape of trajectories immediately before  $t_{kij} - T_{ski} = 0$  for those censored and those experiencing an event suggests that the longitudinal and time-to-event outcomes are related, that an association exists between them. Alternatively, their might be suggestion from healthcare professionals with experience of a given disease area that a particular longitudinal outcome and event are linked, again motivating an analysis exploring the association between the two outcomes. Apart from such methods, the most straightforward way to test for an association would be to fit a joint model, and assess the significance of the association parameter. In the case of assessing the relationship between a longitudinal outcome repeatedly measured over time, and the time until some event, basic assessments of correlation become difficult (for example the multiple measurements per individual over time need to be accounted for). Fitting a joint model provides methods to account for the complex structure of the data, whilst providing an assessment of the relationship between the two outcomes.

As noted in Section 1.6.4, as the joint models investigated involved an unspecified baseline hazard, bootstrap procedures were used to obtain standard errors for parameter estimates [89]. The necessity of bootstrapping increases the time taken to perform a joint analysis, although bootstrapping to obtain standard errors is not unique to joint modelling: for example, using bootstrapping to calculate confidence intervals for longitudinal fits in the *Lme4* package [204] can be time intensive.

It was difficult to extract exact model fitting and bootstrapping times for the simulation studies, due to use of the aforementioned condor system. In this system, analyses often restarted when computers became unavailable, however only a total run time (time analysis first started until time analysis finally completed) was reported. Consequently, samples of 100 models were rerun on a standalone laptop, and the results timed. Bootstrapping procedures were not timed, but a rough calculation of the time they would take to complete can be made by multiplying the time to fit a single model by the number of bootstraps to complete.

For the two-stage MA simulation study, models fitted to low event rate scenarios with association set to 0.5 for homogenous treatment effect data had a mean model fitting time of 6.51 seconds, whilst those with heterogeneous treatment effect had mean fitting time of 6.47 seconds. Models fitted to the high event rate data (for the same level of association) took a mean time of 31.02 seconds for the homogenous treatment effect data, and 31.06 seconds for the heterogeneous treatment effect data. The model fitting times for high event rate data were likely to be larger as the baseline hazard vector will be longer (as it takes weight at each event time [78], it is likely there will be more unique event times in high event rate than low event rate data). Consequently the matrices involved in the model fitting process would be larger, resulting in a slightly higher computational burden to fit the models.

For the one-stage MA simulation study, for scenarios from simulation set 1, (5 studies, a range of association levels), for a sample of 100 of the simulated datasets, mean fitting times were reported at around 40 seconds for model group 1 across the nine scenarios (with slightly faster fitting times for larger association magnitudes). However, model fits for group 4 were much faster at between 10 seconds for scenarios with  $\alpha^{(3)} = 0$  or around 20 seconds for scenarios with  $\alpha^{(3)} > 0$ . With increasing number of included studies, or increasing numbers of individuals within studies, model fitting times should be expected to increase due to the increased size of the supplied dataset.

This simulation study demonstrated that the time required to fit joint meta-analytic models rather than separate time-to-event meta-analytic models is justified where an individual level association exists between the longitudinal and time-to-event outcomes, in order to prevent time-to-event parameter estimates being biased. However, it would be beneficial to accelerate the model fitting process, or the estimation of standard errors, to reduce the time burden of utilising these methods (see Section 8.3.2.2).

#### 7.3.4 Concluding Remarks

The benefit of utilising joint models over standalone time-to-event models has been demonstrated when non-zero association individual level exists between the longitudinal and time-to-event outcomes for both one and two-stage approaches. Increasing study level association was noted to increase the variability in estimates. For the one-stage approach, issues have been identified with models involving study level random effects, and future research has been proposed into models that share only individual level random effects, and contain (but do not share) study level random effects or frailties. Further



details about proposed future research are given in Section 8.3. Overall, these simulation studies have provided in depth investigations of the behaviour of both one and two-stage approaches to meta-analysis of joint longitudinal and time-to-event data.

In the future, I recommend that in a two-stage MA, researchers employ joint models in place of separate longitudinal or time-to-event models to analyse multi-study joint longitudinal and time-to-event data, where evidence of an association between the longitudinal and time-to-event outcomes exists. Association between the outcomes can be evaluated by producing plots of the longitudinal trajectories, stratified by event type (Section 3.2). Similarly, in a one-stage MA, where evidence of association between the longitudinal and time-to-event outcomes exists, a joint modelling approach is employed. This approach should account for between study heterogeneity through fixed interaction terms with study membership, and stratification of baseline hazard. In scenarios similar to those investigated in the simulation studies, joint models that share study level random effects between sub-models should be avoided. However, as separate and joint methods gave similar MA results in simulations and real data analyses when association is insignificant, to minimise resources used separate methods could be justified. The choice of model (association structure, baseline hazard, random effect specification, time-to-event and longitudinal sub-model etc.) should be made based on the requirements of each individual investigation, as one association structure could prove to be more appropriate than others.

## Chapter 8 : Conclusions and Further Work

---

### 8.1 Overview of Thesis and Key Messages

The work presented in this thesis focussed on the development, application and assessment of methods for the meta-analysis of joint longitudinal and time-to-event data. Initial examination of the literature suggested there was little existing research into modelling multi-study joint data. A review of single study joint analyses of medical or biostatistical data (Chapter 2, Sudell et al [93]) suggested that current reporting standards of such analyses could be improved (e.g. by clear specification of the joint model used). This is essential both for the current understanding of the published analysis, but also to ensure that the work could contribute to future AD-MA of joint data. The review suggested that sufficient information to conduct AD-MA were available for the majority of studies reviewed, but that the availability of estimates from each sub-model, and for the association structure, appeared linked to the reason stated for using joint modelling methodology.

Novel methodologies for one and two-stage IPD-MA of joint data were developed (Chapter 3). Given that coefficients from different joint model specifications can have varying interpretations, guidelines were presented and discussed to help to ensure that only parameters with comparable interpretations are quantitatively pooled. The guidelines were designed to provide a structure for two-stage IPD-MA that researchers could follow to produce good quality analyses. A range of methods to account for between study heterogeneity in a one-stage IPD-MA were presented and discussed. Model groups accounted for between study heterogeneity using a mix of fixed interaction terms between covariates of interest and the study membership variable, study level random effects, and stratification of the baseline hazard by study.

The *joinerMeta* package in R was developed during this thesis to provide software to aid researchers analysing multi-study joint data (Chapter 4 and Chapter 5). The package contained functions to extract and meta-analyse coefficients of interest from single study joint model fits, and functions to fit one-stage models to multi-study joint data (with options for interaction terms, study level random effects and stratified baseline hazard). Functionality was also included to simulate and plot multi-study joint data, as well as prepare it in the correct format for analysis. The developmental version of the package is

currently available for download from GitHub, and the fully tested version is available on R's CRAN mirror.

The INDANA dataset [91] was employed to provide a real world example of the application of one and two-stage multi-study joint modelling methodology (Chapter 6). Hypertensive patients were randomised to any drug intervention for hypertension, versus no treatment, placebo or usual care. Pairwise combinations of SBP and each of time to death, time to MI and time to stroke were analysed through univariate joint models. For both the one and two-stage analyses there was little difference between the results from joint or separate models, potentially due to the estimated values of any significant associations being small in magnitude for this particular clinical example. Both one and two-stage analyses consistently estimated that assignment to any drug intervention for hypertension resulted in a significant reduction in SBP. However, in the two-stage analysis few studies displayed significant effect of treatment assignment on risk of MI or death, although more studies observed a significant effect on risk of stroke. In the one-stage analyses, assignment to any drug intervention for hypertension significantly reduced risk of stroke or MI for most analyses conducted (some study specific treatment effects were not statistically significant). However, apart from the STOP [223] trial, no significant effect of treatment assignment on risk of death was observed. It was noted that as these analyses demonstrated methodology rather than identifying covariates linked to the outcome, heterogeneity between included studies in variables known to be linked to cardiovascular disease could lead to heterogeneity in treatment effect and association parameters.

Extensive simulation studies were conducted to assess the performance of the one and two-stage meta-analytic methods (Chapter 7). The two-stage simulation studies compared use of joint models and separate longitudinal or time-to-event models in the first stage of a two-stage MA under a range of scenarios. These included low or high event rates, homogenous or heterogeneous treatment effect between studies, and a range of association parameters. The separate longitudinal and joint approaches produced similar treatment effect estimates across scenarios. However, when non-zero positive association was present, the separate time-to-event analyses underestimated the true treatment effect by an amount that increased as the magnitude of association increased. However, the joint approach estimated the time-to-event sub-model treatment effect well across all scenarios.

The one-stage simulation study applied the models proposed in Chapter 3 to a range of scenarios, including varying individual level and study level association, varying numbers of included studies, and varying between study heterogeneity. Model groups that included study level random effects displayed some issues, especially the group which solely accounted for between study heterogeneity through study level random effects. It was proposed that differing association structures, including those that do not share study level random effects between models, might be more appropriate. The most reliable model of those tested accounted for between study heterogeneity in each sub-model through fixed interaction terms between study membership and covariates of interest, followed by that which included fixed interaction terms in the longitudinal sub-model, and a baseline hazard stratified by study in the time-to-event sub-model.

## 8.2 Implications for research

The research presented in this thesis identifies several points that have implications for future research. Firstly, the use of joint models is increasing, but in single study biostatistical or medical investigations, the reporting of the structure of the joint model used, and of estimates of model parameters of interest, could be improved. This is important, as future AD-MA would be affected by the standard of reporting of joint models in the literature. It is hoped that the research published in Sudell et al [93] will draw attention to, and so improve, this issue.

A range of models for the one-stage analysis of joint longitudinal and time-to-event data were proposed and examined. Several issues with the models were highlighted. The study level association parameter was poorly estimated in model groups involving study level random effects apart from the group that solely represented between study heterogeneity through study level random effects. However, in this group, the remaining model parameters were badly estimated. The best performance came from the model that accounted for between study heterogeneity using fixed interaction terms between the study membership variable and covariates of interest, however this model could become unwieldy as the number of studies included in the meta-analysis increases. It was noted that models that included study level random effects or frailty terms in the longitudinal or time-to-event sub-models may perform better if the study level random effects were not shared between the sub-models, an area noted for future research. As such, this research recommends that, until additional joint meta-analytic models can be examined, multi-study

joint data is modelled using structures that do not share study level random effects between sub-models.

The joint MA models investigated during this thesis present a larger time commitment than standalone longitudinal or time-to-event MA analyses, due to their necessity to bootstrap to obtain standard errors caused by the use of an unspecified baseline hazard in the time-to-event sub-model [89]. However, it has been shown through the one and two-stage simulation studies that this additional time investment is justified when there is a significant association between the longitudinal and time-to-event outcomes. When this occurs, standalone time-to-event MA model underestimate the true time-to-event parameters for non-zero positive associations, whereas joint MA models give estimates closer to the true value. Differences between separate and joint approaches has been shown previously in single study cases [158, 172, 244]; the research in this thesis has demonstrated similar issues for multi-study analyses. Therefore, despite the increased model fitting times caused by multiple included studies, in the presence of association between the longitudinal and time-to-event outcomes, the joint modelling approach remains justified in a meta-analytic setting. The presence of an association can be established by plotting longitudinal trajectories panelled by event type – differences in the trajectories between those experiencing an event and those censored could suggest the presence of an association. The *joinermeta* package developed during this thesis contains plotting functions to accomplish this (Chapter 4).

Both one and two-stage approaches were examined during this thesis. Burke et al [189] have conducted research to identify reasons why the approaches may produce differing results. In this research, a two-stage approach was faster to implement than the one-stage approach for both the real analyses and simulation studies. Bootstrapping to obtain SE was completed with a smaller sample size of  $n_k$  in the two-stage approach (the number of individuals within each study) rather than of  $\sum_{k=1}^K n_k$  in the one-stage approach (the total number of individuals in the meta-dataset), although in the two-stage approach  $K$  separate bootstrap procedures had to complete rather than just one.

A two-stage MA is often sufficient if interest lies solely in estimating treatment effects. Two-methods additionally remove the issue of ecological bias, as they only pool within study information [63, 180, 192]. However, this approach assumes that study estimates can be considered normally distributed, and that their variances are known (an issue for small sample sizes or rare events) [191, 193].

Alternatively, a one-stage approach may be preferred if patient level covariates are to be included in the model. One-stage methods have more exact likelihood specification than two-stage models, and so they circumvent the assumption of normality within studies [190, 191]. However, care must be taken in a one-stage MA to separate within and between study effects [180]. Additionally, as we have also noted during this thesis, one-stage methods tend to be more computationally intensive than two-stage approaches [63, 190].

The choice between one and two-stage approaches is often difficult. Debray et al [194] recommends that, where doubt exists as to the best approach, to plan, conduct, and report the results of both a one and two-stage analysis (a recommendation echoed by Burke et al [189]). It is hoped that the research provided in this thesis will provide methodology, software and guidance for researchers performing meta-analyses of joint longitudinal and time-to-event data in the future.

### 8.3 Planned future work

Research into single study joint models continues to expand. Consequently, a range of meta-analytic joint longitudinal and time-to-event datasets may be available in the future. It is vital that appropriate methods to analyse a range of different scenarios are available, including multivariate multi-study joint models, joint models in a network meta-analytic setting, and incorporation of competing risks options. In addition, developments to the *joinermeta* package would be valuable, including reduction of model fitting times, and expanding the modelling options.

#### 8.3.1 Planned future methodological research

The methodology presented in this thesis are restricted to a univariate joint model with a single longitudinal outcome modelled using a linear mixed effects model, with a single time-to-event outcome modelled using a Cox PH model (which has an unspecified baseline hazard), linked through shared zero mean random effects.

Additionally, examination of the  $W_{2ki}(t)$  in the one-stage INDANA analyses reported in Chapter 6 suggests that it may be beneficial to develop methods to allow the association parameter to vary across studies included in the meta-analysis. Whilst beyond the scope of this thesis, it would increase the flexibility of the joint MA modelling methodology developed here.

Also, it has been noted that the estimates from standalone longitudinal models and joint models have been similar during the real data and simulation study analyses conducted in

this thesis. However, this may not be the case for joint data specifically generated by outcome related dropout in longitudinal studies. Further real data and simulation studies into such dropout data would be beneficial, as it may highlight cases where multi-study longitudinal and joint analyses differ.

However, in reality, many datasets (including the INDANA [91] dataset) contain multiple longitudinal and time-to-event outcomes of interest. Recently, methodology has expanded to allow multiple longitudinal or time-to-event outcomes to be simultaneously modelled in a single study (e.g. Lin et al [245], with software in *joinerML* [181] package). Expanding meta-analytic joint models to allow for multiple longitudinal or time-to-event outcomes would allow a more overall investigation of available data to take place. However, such models would have to link between multiple outcomes, as well as modelling the clustering of individuals within studies. Consequently, it is likely that there will be challenges with model specification and fitting to be reconciled.

Allowing competing risks in the time-to-event outcome to be accounted for in a multi-study joint model would also be beneficial. An event such as death may occur for several reasons. Competing risks models allows the risk of the event due to each reason to be modelled. Joint models allowing for competing risks in single study cases exists in the literature (e.g. [246]). However, work to expand this methodology to allow for datasets containing additional levels of clustering has not been undertaken.

Network meta-analyses expand standard meta-analyses by allowing a group of studies examining the same disease area, but comparing different combinations of possible treatments, to collectively assess and compare different treatment combinations. A recent example of a network MA of time-to-event epilepsy data is given in Nevitt et al [247]. Information comparing treatment options can stem from direct comparisons within studies, or from indirect information supplied through the network (e.g. an assessment of treatment A versus C can be made using studies comparing treatment A versus treatment B, and treatment B versus treatment C). Utilisation of joint models in this area would be intriguing, as the network of information must allow for the structure (sub-models linked through an association structure) of the joint model, resulting in linked networks for longitudinal, time-to-event and association coefficients.

An additional area of interest would be to examine the feasibility of combining estimates from separate longitudinal or time-to-event models, and joint models in the second stage of a two-stage IPD-MA, or in an AD-MA. This question requires careful consideration, as

the requirement of availability of both longitudinal and time-to-event data in a joint MA may restrict the studies able to contribute to the meta-analysis. However, this thesis has demonstrated that in a meta-analysis, using estimates from separate time-to-event models results in a pooled value that underestimates the true value when significant association exists between the longitudinal and time-to-event outcomes. Conversely, the estimates from joint and separate longitudinal models were similar regardless of the significance or magnitude of association. Investigation of the benefit of using both separate and joint model results may allow expansion of the data that can contribute to meta-analyses.

Dynamic prediction has been an area of increasing research within joint modelling [134, 163, 248-250], where predictions about an individual's future risk of an event can be updated as they provide additional longitudinal measurements. It would be beneficial to investigate dynamic prediction from a one-stage IPD-MA joint model, as such a prediction model would take into account all available information on a disease area, rather than just that proceeding from a single study. Such research would help to maximise the benefit of IPD-MA joint investigations, whilst providing patients with more accurate risk predictions.

### 8.3.2 Planned developments to the *joinerMeta* package

The *joinerMeta* package has been submitted to CRAN in order that it can be downloaded directly from R rather than through GitHub. This should increase uptake of use of the provided code. As well as ensuring the package is easily accessible, there are a range of areas for which extension of the *joinerMeta* package would be beneficial.

#### 8.3.2.1 Extensions to functions for two-stage MA

Firstly, for the *jointmeta2* function, it would be beneficial to extend the code to allow joint modelling fits from other packages to be supplied. Whilst the one-stage function *jointmeta1* in the *joinerMeta* package is designed to handle multi-study data, it can handle single study data providing the data contains a study membership variable (which would take a common value across all individuals). As such, it would be useful to allow the *jointmeta2* function to accept single study joint model fits fitted using *jointmeta1*. In addition, the recent *joinerML* package allows multivariate (multiple longitudinal outcome) as well as univariate (single longitudinal outcome) joint models to be fitted. Expanding *jointmeta2* to take model fits from a wider range of packages would be beneficial, as it would increase the applicability of the package.

In addition, in Section 3.4.1 it was noted that performing multivariate meta-analyses of joint model fits, rather than separate meta-analyses for each parameter of interest could



be beneficial in that it will model the correlation between parameters [35, 38, 40]. In the future, it may be beneficial to include the option to perform multivariate MA in the *jointmeta2* function. This would rely on the covariance between parameters to be obtainable from the joint model fits. With this extension in mind, the covariance matrix for model parameters is returned from the bootstrapping process in *jointmeta1* fits. In addition, the Hessian can be extracted from fits from the *JM* package. This extension would improve the quality of two-stage MA conducted using *joinermeta*.

### 8.3.2.2 Extensions to functions for one-stage MA

Currently, the one-stage function *jointmeta1* employs a pseudo-adaptive Gaussian quadrature procedure to integrate over random effects included in the model [196]. Although more computationally intensive, a fully adaptive procedure has the potential (through relocation of the quadrature points at each iteration) to reduce model fitting times. Adaptive Gaussian quadrature procedures for three level structures with multi-level random effects were discussed by Rabe-Hesketh et al [212]. They note that it is necessary to rescale multi-level random effects during the estimation procedure. Some initial work was undertaken during this thesis to implement this procedure, however further time was needed to interpret and implement the procedure provided in the paper. In the future, implementation of this procedure as an alternative to the pseudo-adaptive procedure for the *jointmeta1* function is planned.

Additionally, the function currently employs the EM algorithm [197] to fit the model. Various modifications to this method (which could improve model fitting times) could be made. Vermunt [251] noted that the computation time of the EM algorithm increases exponentially with the number of level 2 units within each level 3 unit. This was not experienced during our research, as the iterative process was exploited to hold estimates at one level constant when estimating functions of random effects at the other level, however with implementation of fully adaptive Gaussian quadrature, this will be an issue. Vermunt [251] demonstrates implementation of an edited EM algorithm designed for use with parametric or non-parametric hierarchical non-linear models with more than two levels, which could be implemented to solve this issue. Another alternative is to employ an EM algorithm with a Monte-Carlo step as in the *joinerML* package [181].

Currently, bootstrapping to obtain standard errors is time consuming. Lin et al [245] have suggested that the underestimation noted by Hsieh et al [89] may not be so severe as to prevent use of an approximate standard error estimation procedure during model fitting, although bootstrapping to estimate final standard errors may be advisable. Some initial

work was performed when writing the *joinerMeta* package to calculate such approximate standard errors for joint MA models, however the results for study level random effects appeared unreliable. It may be that the approximate standard error procedure provides better estimations when study level random effects are not involved in the association structure. In the future, the approximate standard errors approach for joint models employing an unspecified baseline hazard will be re-examined and implemented in the *jointmeta1* function.

The *jointmeta1* function could be extended to fit joint models with varying association structures. Association structures each have different interpretations, and their selection should be driven by the research question, not the availability of software. Expansions could include the current value sharing structure (which inserts the entire longitudinal trajectory without the error term into the time-to-event sub-model) or the current slope structure (first derivative of the longitudinal trajectory with respect to time). Additionally, options to just share the zero mean random effects at one level rather than both may be beneficial, as discussed in Chapter 7. Additionally, allowing separate association parameters for each shared random effect would allow more in-depth assessment of the effect of each random effect on the risk of an event. Allowing structures that link the history of the longitudinal outcome to the time-to-event outcome would also be useful [52, 75].

Allowing alternative baseline hazard specifications could also be beneficial. Permitting parametric baseline hazards or spline based baseline hazards (such as in the single study joint modelling package *JM* [81]) could result in fewer coefficients being estimated, and would remove the necessity to bootstrap to determine SE, thus reducing model fitting times.

Expanding the types of sub-models permitted in the package would also be useful. Many continuous longitudinal outcomes are non-linear, and so including longitudinal sub-model options such as use of splines would improve the applicability of the package. In addition, permitting alternative covariance structures for the error term (e.g. auto-regressive or completely unspecified rather than independently and identically distributed) may allow more appropriate models to be fitted.

In cases where a PH model is not appropriate for the time-to-event outcome, an alternative such as the AFT model might be preferred. A range of modelling options have been used

for the MA of time-to-event data [58, 59], which could be incorporated into the time-to-event sub-model of a multi-study joint model in the *jointmeta1* function.

### 8.3.2.3 Other extensions to the package

Additionally, extensions to the functions to produce plots of the study specific data would be beneficial. For example, adjusting the plotting function such that confidence bands around the survival curves can be easily included may increase the clarity of the study specific Kaplan-Meier plots. It would also be beneficial to include a function to easily produce visualisations of the baseline hazards of analyses.

### 8.3.3 Planned future applied investigations

During this thesis the methods developed were applied to the INDANA dataset of hypertensive patients [91]. This analysis was beneficial to test the methods in a real world setting, and to motivate subsequent simulation investigations. However, application of the methods to data from other clinical areas may identify additional avenues of research, as well as informing clinical practice. For example, other disease areas could present data with different event rates, number of longitudinal measurements, levels of between study or between individual heterogeneity, or levels of association, or other characteristics, which that could prompt development of new methodology. As the review of reporting quality in single study joint analyses reported (Chapter 2, Sudell et al [93]), joint models have been applied in a wide range of clinical areas, including cancer, HIV/AIDS, transplants, and cognitive decline. As such, it is conceivable that a wider range of joint multi-study datasets will be established in the future.

## 8.4 Conclusion

In conclusion, the research presented in this thesis has examined the meta-analysis of multi-study joint longitudinal and time-to-event data in a frequentist setting. A range of topics have been investigated, including current reporting standards of single study joint models with a view to future meta-analyses, one and two-stage methods for meta-analysis of joint data and their behaviour under different scenarios, and an R package containing code useful when implementing a joint data meta-analysis. It has been shown that joint approaches are preferable to separate approaches to avoid underestimation of time-to-event coefficients for both one and two-stage joint data MA where association is present between the longitudinal and time-to-event outcomes. There remain a wide range of areas where future research would be beneficial.

## Bibliography

---

1. Glass, G.V., *Primary, Secondary, and Meta-Analysis of Research*. 1976, American Educational Research Association. p. 3.
2. Whitehead, A., *Meta-analysis of controlled clinical trials / Anne Whitehead*. Statistics in practice. 2002: Chichester : John Wiley & Sons, 2002.
3. Riley, R.D., M.C. Simmonds, and M.P. Look, *Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods*. JOURNAL OF CLINICAL EPIDEMIOLOGY, 2007. **60**(5): p. 431-439.
4. Sutton, A.J., D. Kendrick, and C.A.C. Coupland, *Meta-analysis of individual- and aggregate-level data*. STATISTICS IN MEDICINE, 2008. **27**: p. 651-669.
5. Donegan, S., et al., *Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: Individual patient data may be beneficial if only for a subset of trials*. STATISTICS IN MEDICINE, 2012. **31**(29): p. 3840-3857.
6. Lambert, P.C., et al., *A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis*. Journal of Clinical Epidemiology, 2002. **55**(1): p. 86-94.
7. *Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.*
8. Higgins, J.P.T., I.R. White, and J. Anzures-Cabrera, *Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales*.
9. Berlin, J.A., et al., *Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head*. STATISTICS IN MEDICINE, 2002. **21**(3): p. 371-387.
10. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials*. Controlled Clinical Trials, 1986. **7**(3): p. 177-188.
11. Thompson, S.G. and J.P.T. Higgins, *How should meta-regression analyses be undertaken and interpreted?* Statistics in Medicine, 2002. **21**(11): p. 1559-1573.
12. Chalmers, I., *The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care*. Annals Of The New York Academy Of Sciences 1993. **703**: p. 156-63.
13. Stewart, L.A. and M.K.B. Parmar, *Meta-analysis of the literature or of individual patient data: Is there a difference?* Lancet, 1993. **341**(8842): p. 418.
14. Nevitt, S.J., et al., *Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review*. BMJ, 2017. **357**.
15. Simmonds, M.C., et al., *Meta-analysis of individual patient data from randomized trials: a review of methods used in practice*. Clinical Trials (London, England), 2005. **2**(3): p. 209-217.
16. Stewart LA, Tierney JF, Clarke M. Chapter 19: Reviews of individual patient data. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
17. Stewart, L.A. and J.F. Tierney, *To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data*. Evaluation & the Health Professions, 2002. **25**(1): p. 76-97.
18. Tudur Smith, C., et al., *Individual participant data meta-analyses compared with meta-analyses based on aggregate data*. Cochrane Database of Systematic Reviews, 2016(9).

19. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
20. Higgins, J.P.T. and S.G. Thompson, Quantifying heterogeneity in a meta-analysis.
21. Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
22. Whitehead, A., et al., *Meta-analysis of ordinal outcomes using individual patient data*.
23. Riley, R.D., et al., *Meta-analysis of continuous outcomes combining individual patient data and aggregate data*. *Statistics in Medicine*, 2008. **27**(11): p. 1870-1893.
24. Curtin, F., D.G. Altman, and D. Elbourne, *Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes*.
25. Curtin, F., D. Elbourne, and D.G. Altman, *Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes*.
26. Curtin, F., D. Elbourne, and D.G. Altman, *Meta-analysis combining parallel and cross-over clinical trials. III: The issue of carry-over*.
27. Weiss, R.E., *Modeling Longitudinal Data*. 1 ed. Springer Texts in Statistics. 2005, United States: Springer. 420.
28. Diggle, P., et al., *Analysis of Longitudinal Data*. 2 ed. Oxford Statistical Science Series, ed. A.C. Atkinson, et al. 2002, United States: Oxford University Press. 379.
29. Laird, N.M. and J.H. Ware, *Random-Effects Models for Longitudinal Data*. 1982, Biometric Society. p. 963.
30. Henderson, R., P. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data*. *Biostatistics (Oxford, England)*, 2000. **1**(4): p. 465-480.
31. Jones, A.P., et al., *Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials [corrected] [published erratum appears in CLIN TRIALS 2009;6(3):288]*. *Clinical Trials*, 2009. **6**(1): p. 16-27.
32. Ishak, K.J., et al., *Meta-analysis of longitudinal studies*. *Clinical Trials*, 2007. **4**(5): p. 525-539.
33. Thompson, S., et al., *A proposed method of bias adjustment for meta-analyses of published observational studies*. *International Journal of Epidemiology*, 2011. **40**(3): p. 765-777.
34. Gurrin, L.C. and L. Turkovic, *Combining Individual Participant Data and Summary Statistics from both Continuously Valued and Binary Variables to Estimate Regression Parameters*. *Australian & New Zealand Journal of Statistics*, 2012. **54**(1): p. 1-21.
35. Jackson, D., R. Riley, and I.R. White, *Multivariate meta-analysis: potential and promise*. *Statistics in Medicine*, 2011. **30**(20): p. 2481-2498.
36. Becker, B.J., *Multivariate Meta-Analysis*, in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. 2000, Elsevier. p. 499-525.
37. Berkey, C.S., et al., *Meta-analysis of multiple outcomes by regression with random effects*. *Statistics in Medicine*, 1998. **17**(22): p. 2537-2550.
38. Riley, R.D., *Multivariate meta-analysis: the effect of ignoring within-study correlation*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2009. **172**(4): p. 789-811.

39. Riley, R.D., et al., *An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes*. *Statistics In Medicine*, 2007. **26**(1): p. 78-97.
40. Riley, R.D., et al., *Multivariate meta-analysis using individual participant data*. *Research Synthesis Methods*, 2015. **6**(2): p. 157-174.
41. van Houwelingen, H.C., L.R. Arends, and T. Stijnen, *Advanced methods in meta-analysis: multivariate approach and meta-regression*. *Stat Med*, 2002. **21**(4): p. 589-624.
42. Rubin, D.B., *Inference and Missing Data*. *Biometrika*, 1976(3): p. 581.
43. Little, R.J.A. and D.B. Rubin, *Statistical analysis with missing data*. [electronic boo. Online access with subscription: Proquest Ebook Central. 2002: Second edition.
44. Maas, C.J.M., J.J. Hox, and G.J.L.M. Lensvelt-Mulders, *Longitudinal Meta-analysis. Quality & Quantity*, 2004. **38**(4): p. 381-389.
45. Peters, J.L. and K.L. Mengersen, *Meta-analysis of repeated measures study designs*. *Journal of Evaluation in Clinical Practice*, 2008. **14**(5): p. 941-950.
46. Ahn, J.E. and J.L. French, *Longitudinal aggregate data model-based meta-analysis with NONMEM: Approaches to handling within treatment arm correlation*. *Journal of Pharmacokinetics and Pharmacodynamics*, 2010. **37**(2): p. 179-201.
47. Dakin, H.A., et al., *Mixed treatment comparison of repeated measurements of a continuous endpoint: An example using topical treatments for primary open-angle glaucoma and ocular hypertension*. *Statistics in Medicine*, 2011. **30**(20): p. 2511-2535.
48. Boucher, M., *Imputation of missing variance data using non-linear mixed effects modelling to enable an inverse variance weighted meta-analysis of summary-level longitudinal data: A case study*. *Pharmaceutical Statistics*, 2012. **11**(4): p. 318-324.
49. Collett, D., *Modelling Survival Data in Medical Research*. 2 ed. *Texts in Statistical Science Series*, ed. C. Chatfield, M. Tanner, and J. Zidek. Vol. 1. 2003: Chapman & Hall/CRC. 391.
50. Cox, D.R., *Regression Models and Life-Tables*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1972. **34**(2): p. 187-220.
51. Kalbfleisch, J. and R. Prentice, *The Statistical Analysis of Failure Time Data*. 2 ed. 2002, New York: Wiley.
52. Rizopoulos, D., *Joint Models for Longitudinal and Time-to-Event Data With Applications in R*. 1 ed. *Chapman and Hall/CRC Biostatistics Series*, ed. S. Chow, et al. Vol. 1. 2012. 261.
53. Andersen, P.K. and R.D. Gill, *Cox's Regression Model for Counting Processes: A Large Sample Study*. *The Annals of Statistics*, 1982(4): p. 1100.
54. Fleming, T.R. and D.P. Harrington, *Nonparametric estimation of the survival distribution in censored data*. *Communications in Statistics - Theory and Methods*, 1984. **13**(20): p. 2469-2486.
55. Andersen, P., et al., *Statistical Models Based on Counting Processes*. 1993: Springer-Verlag, New York.
56. Tudur Smith, C., P.R. Williamson, and A.G. Marson, *Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes*. *Statistics in Medicine*, 2005. **24**(9): p. 1307-1319.
57. Tudur Smith, C., *Individual patient data meta-analysis with time-to-event outcomes*. 2004: Liverpool : Thesis Ph.D., 2004.
58. Crowther, M.J., M.P. Look, and R.D. Riley, *Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis*. *Statistics in Medicine*, 2014. **33**(22): p. 3844.

59. Crowther, M.J., et al., *Individual patient data meta-analysis of survival data using Poisson regression models*. BMC MEDICAL RESEARCH METHODOLOGY, 2012. **12**.
60. Katsahian, S., et al., *Practical methodology of meta-analysis of individual patient data using a survival outcome*. Contemporary Clinical Trials, 2008. **29**(2): p. 220-230.
61. Michiels, S., et al., *Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses*. Journal of Clinical Epidemiology, 2005. **58**(3): p. 238-245.
62. Rondeau, V., et al., *Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach*. Stat Med, 2008. **27**: p. 1894–1910.
63. Thompson, S., et al., *Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies*. International Journal of Epidemiology, 2010. **39**(5): p. 1345-1359.
64. Parmar, M.K., V. Torri, and L. Stewart, *Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints*. Statistics In Medicine, 1998. **17**(24): p. 2815-2834.
65. Tudur, C., et al., *The value of the aggregate data approach in meta-analysis with time-to-event outcomes*. Journal of the Royal Statistical Society. Series A: Statistics in Society, 2001. **164**(2): p. 357-370.
66. Tierney, J.F., et al., *Practical methods for incorporating summary time-to-event data into meta-analysis*. Trials, 2007. **8**.
67. Williamson, P.R., et al., *Aggregate data meta-analysis with time-to-event outcomes*. Statistics In Medicine, 2002. **21**(22): p. 3337-3351.
68. Duchateau, L., et al., *Estimating number of events from the Kaplan-Meier curve for incorporation in a literature-based meta-analysis: what you don't see you can't get!* Biometrics, 2000. **56**(3): p. 886-892.
69. Arends, L.R., M.G.M. Hunink, and T. Stijnen, *Meta-analysis of summary survival curve data*. Statistics in Medicine, 2008. **27**(22): p. 4381-4396.
70. Bennett, M.M., et al., *Comparison of Bayesian and Frequentist Meta-Analytical Approaches for Analyzing Time to Event Data*. Journal of Biopharmaceutical Statistics, 2013. **23**(1): p. 129-145.
71. Duchateau, L., et al., *Individual patient-versus literature-based meta-analysis of survival data: Time to event and event rate at a particular time can make a difference, an example based on head and neck cancer*. CONTROLLED CLINICAL TRIALS, 2001. **22**(5): p. 538-547.
72. Barrett, J.K., et al., *Two-stage meta-analysis of survival data from individual participants using percentile ratios*. Statistics in Medicine, 2012. **31**(30): p. 4296-4308.
73. Siannis, F., et al., *One-stage parametric meta-analysis of time-to-event outcomes*. Statistics In Medicine, 2010. **29**(29): p. 3030-3045.
74. Davidian, M., et al., *Discussion of joint modeling longitudinal and survival data*. Statistica Sinica, 2004. **14**(3): p. 621-624.
75. Gould, A.L., et al., *Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group*. Stat Med, 2015. **34**(14): p. 2181-95.
76. Proust-Lima, C., et al., *Joint latent class models for longitudinal and time-to-event data: A review*. Statistical Methods in Medical Research, 2014. **23**(1): p. 74-90.
77. Elashoff, R.M., G. Li, and N. Li, *Joint Modeling of Longitudinal and Time-to-Event Data*. 2017: CRC Press.

78. Wulfsohn, M.S. and A.A. Tsiatis, *A Joint Model for Survival and Longitudinal Data Measured with Error*. International Biometric Society, 1997(1): p. 330.
79. R Core Team, *R: A Language and Environment for Statistical Computing*. 2015, R Foundation for Statistical Computing: Vienna, Austria.
80. Philipson, P., et al., *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.2.2, <https://github.com/graemeleehickey/joineR/>. . 2017.
81. Rizopoulos, D., *JM: An R package for the joint modelling of longitudinal and time-to-event data*. Journal of Statistical Software, 2010. **35**(9): p. 1-33.
82. Rizopoulos, D., *The R Package Jmbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data using MCMC*. 2014.
83. Crowther, M.J., *STJM: Stata module to fit shared parameter joint models of longitudinal and survival data*. 2013.
84. StataCorp. *Stata. Statistical Software*. College Station, TX: StataCorp LP.
85. SAS software, Copyright © 2017 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
86. Rizopoulos, D., *Joint Modeling of Longitudinal and Time-to-Event Data: Challenges and Future Directions*. Advances in Theoretical & Applied Statistics, 2013: p. 199.
87. Ibrahim, J.G., H. Chu, and L.M. Chen, *Basic concepts and methods for joint models of longitudinal and survival data*. Journal of Clinical Oncology, 2010. **28**(16): p. 2796-2801.
88. Patiwan, Y., *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. 2001: Claredon Press, Oxford.
89. Hsieh, F., Y.-K. Tseng, and J.-L. Wang, *Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited*. Biometrics, 2006. **62**(4): p. 1037-1043.
90. Brombin, C., C. Di Serio, and P.M.V. Rancoita, *Joint modeling of HIV data in multicenter observational studies: A comparison among different approaches*. Statistical Methods In Medical Research, 2016. **25**(6): p. 2472-2487.
91. Gueyffier, F., et al., *INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results*. Therapie, 1995. **50**: p. 353-362.
92. Tsiatis, A.A. and M. Davidian, *Joint modeling of longitudinal and time-to-event data: An overview*. STATISTICA SINICA, 2004. **14**(3): p. 809-834.
93. Sudell, M., R. Kolamunnage-Dona, and C. Tudur-Smith, *Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis*. BMC Medical Research Methodology, 2016. **16**(168).
94. Andrinopoulou, E.R., et al., *An introduction to mixed models and joint modeling: Analysis of valve function over time*. Annals of Thoracic Surgery, 2012. **93**(6): p. 1765-1772.
95. Andrinopoulou, E.R., et al., *Joint modeling of two longitudinal outcomes and competing risk data*. Statistics in Medicine, 2014. **33**(18): p. 3167-3178.
96. Argyropoulos, C., et al., *Dialyzer reuse and outcomes of high flux dialysis*. PLoS ONE, 2015. **10**(6).
97. Artaud, F., et al., *Decline in fast gait speed as a predictor of disability in older adults*. Journal of the American Geriatrics Society, 2015. **63**(6): p. 1129-1136.
98. Berzuini, C. and C. Allemani, *Effectiveness of potent antiretroviral therapy on progression of human immunodeficiency virus: Bayesian modelling and model checking via counterfactual replicates*. Journal of the Royal Statistical Society. Series C: Applied Statistics, 2004. **53**(4): p. 633-650.
99. Cella, D., et al., *Survival-adjusted health-related quality of life (HRQL) among patients with metastatic breast cancer receiving paclitaxel plus bevacizumab versus*



- paclitaxel alone: Results from Eastern Cooperative Oncology Group Study 2100 (E2100)*. Breast Cancer Research and Treatment, 2011. **130**(3): p. 855-861.
100. Chang, C.C.H., et al., *Minimizing attrition bias: A longitudinal study of depressive symptoms in an elderly cohort*. International Psychogeriatrics, 2009. **21**(5): p. 869-878.
  101. Chen, Q., et al., *Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates*. Statistics in Medicine, 2014. **33**(26): p. 4560-4576.
  102. Daher Abdi, Z., et al., *Impact of longitudinal exposure to mycophenolic acid on acute rejection in renal-transplant recipients using a joint modeling approach*. Pharmacological Research, 2013. **72**: p. 52-60.
  103. Dantan, E., et al., *Joint model with latent state for longitudinal and multistate data*. Biostatistics, 2011. **12**(4): p. 723-736.
  104. Deslandes, E. and S. Chevret, *Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: Application to ICU data*. BMC Medical Research Methodology, 2010. **10**.
  105. Du, H., E.A. Hahn, and D. Cella, *The impact of missing data on estimation of health-related quality of life outcomes: An analysis of a randomized longitudinal clinical trial*. Health Services and Outcomes Research Methodology, 2011. **11**(3-4): p. 134-144.
  106. Duvignac, J., et al., *CD4+ T-lymphocytes natural decrease in HAART-naïve HIV-infected adults in Abidjan*. HIV Clinical Trials, 2008. **9**(1): p. 26-35.
  107. Ediebah, D.E., et al., *Joint modeling of longitudinal health-related quality of life data and survival*. Quality of Life Research, 2015. **24**(4): p. 795-804.
  108. Elhaj, M., et al., *Can serum surfactant protein D or CC-chemokine ligand 18 predict outcome of interstitial lung disease in patients with early systemic sclerosis?* Journal of Rheumatology, 2013. **40**(7): p. 1114-20.
  109. Fairclough, D.L., et al., *Handling missing quality of life data in HIV clinical trials: what is practical?* Qual Life Res, 2008. **17**(1): p. 61-73.
  110. Garre, F.G., et al., *A joint latent class changepoint model to improve the prediction of time to graft failure*. Journal of the Royal Statistical Society. Series A: Statistics in Society, 2008. **171**(1): p. 299-308.
  111. Ghisletta, P., *Application of a joint multivariate longitudinal-survival analysis to examine the terminal decline hypothesis in the Swiss Interdisciplinary Longitudinal Study on the Oldest Old*. Journals of Gerontology - Series B Psychological Sciences and Social Sciences, 2008. **63**(3): p. P185-92.
  112. Ghisletta, P., J.J. McArdle, and U. Lindenberger, *Longitudinal cognition-survival relations in old and very old age: 13-year data from the Berlin aging study*. European Psychologist, 2006. **11**(3): p. 204-223.
  113. Goyal, N. and R. Gomeni, *A latent variable approach in simultaneous modeling of longitudinal and dropout data in schizophrenia trials*. European Neuropsychopharmacology, 2013. **23**(11): p. 1570-1576.
  114. Graham, P.L., L.M. Ryan, and M.A. Luszcz, *Joint modelling of survival and cognitive decline in the Australian Longitudinal Study of Ageing*. Journal of the Royal Statistical Society. Series C: Applied Statistics, 2011. **60**(2): p. 221-238.
  115. Guler, I., et al., *Joint modelling for longitudinal and time-to-event data: Application to liver transplantation data*, in *14th International Conference on Computational Science and Its Applications, ICCSA 2014*. 2014, Springer Verlag: Guimaraes. p. 580-593.

116. Husser, O., et al., *Tumor marker carbohydrate antigen 125 predicts adverse outcome after transcatheter aortic valve implantation*. JACC Cardiovasc Interv, 2013. **6**(5): p. 487-96.
117. Ibrahim, J.G., M.H. Chen, and D. Sinha, *Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials*. Statistica Sinica, 2004. **14**(3): p. 863-883.
118. Jacoby, A., et al., *Quality-of-life outcomes of initiating treatment with standard and newer antiepileptic drugs in adults with new-onset epilepsy: Findings from the SANAD trial*. Epilepsia, 2015. **56**(3): p. 460-472.
119. Khoundabi, B., et al., *Acute Kidney Injury in ICU Patients Following Non-Cardiac Surgery at Masih Daneshvari Hospital: Joint Modeling Application*. Tanaffos, 2015. **14**(1): p. 49-54.
120. Kupferman, J.C., et al., *BP control and left ventricular hypertrophy regression in children with CKD*. J Am Soc Nephrol, 2014. **25**(1): p. 167-74.
121. Kypriotakis, G., et al., *The longitudinal relationship between quality of life and survival in advanced stage cancer*. Psycho-Oncology, 2015.
122. Lawson, A.B., et al., *Bayesian latent structure modeling of walking behavior in a physical activity intervention*. Stat Methods Med Res, 2014.
123. Levine, S.Z., et al., *Joint modeling of dropout and outcome in three pivotal clinical trials of schizophrenia*. Schizophrenia Research, 2015. **164**(1-3): p. 122-126.
124. Liu, X., et al., *Does C-reactive protein predict the long-term progression of interstitial lung disease and survival in patients with early systemic sclerosis?* Arthritis Care and Research, 2013. **65**(8): p. 1375-1380.
125. Malehi, A.S., et al., *Assessing the autoantibody levels in relation to recurrence of pemphigus: Joint modeling of longitudinal measurements and recurrent event times*. Iranian Red Crescent Medical Journal, 2014. **16**(2).
126. Marioni, R.E., et al., *Cognitive lifestyle jointly predicts longitudinal cognitive decline and mortality risk*. European Journal of Epidemiology, 2014. **29**(3): p. 211-219.
127. Mayeda, E.R., et al., *Does Type 2 Diabetes Increase Rate of Cognitive Decline in Older Mexican Americans?* Alzheimer Disease and Associated Disorders, 2015.
128. McLain, A.C., R. Sundaram, and G.M. Buck Louis, *Joint analysis of longitudinal and survival data measured on nested timescales by using shared parameter models: An application to fecundity data*. Journal of the Royal Statistical Society. Series C: Applied Statistics, 2015. **64**(2): p. 339-357.
129. Medeiros, F.A., et al., *Evaluation of progressive neuroretinal rim loss as a surrogate end point for development of visual field loss in glaucoma*. Ophthalmology, 2014. **121**(1): p. 100-109.
130. Meira-Freitas, D., et al., *Predicting progression in glaucoma suspects with longitudinal estimates of retinal ganglion cell counts*. Investigative Ophthalmology and Visual Science, 2013. **54**(6): p. 4174-4183.
131. Meira-Freitas, D., et al., *Predicting progression of glaucoma from rates of frequency doubling technology perimetry change*. Ophthalmology, 2014. **121**(2): p. 498-507.
132. Miki, A., et al., *Rates of retinal nerve fiber layer thinning in glaucoma suspect eyes*. Ophthalmology, 2014. **121**(7): p. 1350-1358.
133. Murphy, T.E., et al., *A longitudinal, observational study with many repeated measures demonstrated improved precision of individual survival curves using bayesian joint modeling of disability and survival*. Experimental Aging Research, 2015. **41**(3): p. 221-239.
134. Njagi, E.N., et al., *A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients*. Statistical Modelling: An International Journal, 2013. **13**(3): p. 179.

135. Núñez, J., et al., *Red blood cell distribution width is longitudinally associated with mortality and anemia in heart failure patients*. *Circulation Journal*, 2014. **78**(2): p. 410-418.
136. Pan, J., et al., *Joint longitudinal and survival-cure models in tumour xenograft experiments*. *Statistics in Medicine*, 2014. **33**(18): p. 3229-3240.
137. Pauler, D.K. and D.M. Finkelstein, *Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes*. *Statistics in Medicine*, 2002. **21**(24): p. 3897-3911.
138. Paulsen, J.S., et al., *Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study*. *Lancet Neurology*, 2014. **13**(12): p. 1193-201.
139. Pike, F., L.A. Weissfeld, and C.C.H. Chang, *Joint modeling of multivariate censored longitudinal and event time data with application to the Genetic Markers of Inflammation Study*. *Journal of Applied Statistics*, 2014. **41**(10): p. 2178-2191.
140. Pilla Reddy, V., et al., *Modelling and simulation of the positive and negative syndrome scale (PANSS) time course and dropout hazard in placebo Arms of schizophrenia clinical trials*. *Clinical Pharmacokinetics*, 2012. **51**(4): p. 261-275.
141. Sargent-Cox, K.A., K.J. Anstey, and M.A. Luszcz, *Longitudinal change of self-perceptions of aging and mortality*. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 2014. **69**(2): p. 168-173.
142. Seid, A., et al., *Joint modeling of longitudinal CD4 cell counts and time-to-default from HAART treatment: A comparison of separate and joint models*. *Electronic Journal of Applied Statistical Analysis*, 2014. **7**(2): p. 292-314.
143. Shen, Y., et al., *Joint modeling tumor burden and time to event data in oncology trials*. *Pharmaceutical Statistics*, 2014. **13**(5): p. 286-293.
144. Skupien, J., et al., *Synergism between circulating tumor necrosis factor receptor 2 and HbA(1c) in determining renal decline during 5-18 years of follow-up in patients with type 1 diabetes and proteinuria*. *Diabetes Care*, 2014. **37**(9): p. 2601-8.
145. Su, C.T., et al., *Changes in anthropometry and mortality in maintenance hemodialysis patients in the HEMO study*. *American Journal of Kidney Diseases*, 2013. **62**(6): p. 1141-1150.
146. Tang, X., et al., *Increasing Chimerism after Allogeneic Stem Cell Transplantation Is Associated with Longer Survival Time*. *Biology of Blood and Marrow Transplantation*, 2014. **20**(8): p. 1139-1144.
147. Terrin, N., A.M. Rodday, and S.K. Parsons, *Joint models for predicting transplant-related mortality from quality of life data*. *Quality of Life Research*, 2015. **24**(1): p. 31-39.
148. Thabut, G., et al., *Survival benefit of lung transplant for cystic fibrosis since lung allocation score implementation*. *American Journal of Respiratory & Critical Care Medicine*, 2013. **187**(12): p. 1335-40.
149. Thiebaut, R., et al., *Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection*. *Statistics in Medicine*, 2005. **24**(1): p. 65-82.
150. van den Hout, A. and G. Muniz-Terrera, *Joint models for discrete longitudinal outcomes in aging research*. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 2015.
151. Van Der Linde, D., et al., *Natural history of discrete subaortic stenosis in adults: A multicentre study*. *European Heart Journal*, 2013. **34**(21): p. 1548-1556.

152. Wang, Y. and J.M.G. Taylor, *Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome*. Journal of the American Statistical Association, 2001. **96**(455): p. 895-905.
153. Watanabe, H., et al., *Factors affecting longitudinal functional decline and survival in amyotrophic lateral sclerosis patients*. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 2015. **16**(3-4): p. 230-236.
154. Wu, L., W. Liu, and X.J. Hu, *Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors*. Biometrics, 2010. **66**(2): p. 327-335.
155. Xue, Q.L., et al., *Heterogeneity in rate of decline in grip, hip, and knee strength and the risk of all-cause mortality: The women's health and aging study II*. Journal of the American Geriatrics Society, 2010. **58**(11): p. 2076-2084.
156. Yu, M., J.M.G. Taylor, and H.M. Sandler, *Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model*. Journal of the American Statistical Association, 2008. **103**(481): p. 178-187.
157. Zhang, J.P., et al., *Joint Modeling of Longitudinal Changes in Depressive Symptoms and Mortality in a Sample of Community-Dwelling Elderly People*. PSYCHOSOMATIC MEDICINE, 2009. **71**(7): p. 704-714.
158. Guo, X. and B.P. Carlin, *Separate and joint modeling of longitudinal and event time data using standard computer packages*. AMERICAN STATISTICIAN, 2004. **58**(1): p. 16-24.
159. Crowther, M.J., K.R. Abrams, and P.C. Lambert, *Flexible parametric joint modelling of longitudinal and survival data*. Statistics in Medicine, 2012. **31**(30): p. 4456-4471.
160. Crowther, M.J., K.R. Abrams, and P.C. Lambert, *Joint modeling of longitudinal and survival data*. Stata Journal, 2013. **13**(1): p. 165-184.
161. Diggle, P.J., I. Sousa, and A.G. Chetwynd, *Joint modelling of repeated measurements and time-to-event outcomes: The fourth Armitage lecture*. STATISTICS IN MEDICINE, 2008. **27**(16): p. 2981-2998.
162. Proust-Lima, C., et al., *Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach*. Computational Statistics and Data Analysis, 2009. **53**(4): p. 1142-1154.
163. Rizopoulos, D., *Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data*. Biometrics, 2011. **67**(3): p. 819-829.
164. Rizopoulos, D. and P. Ghosh, *A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event*. Statistics in Medicine, 2011. **30**(12): p. 1366-1380.
165. Rondeau, V., Y. Mazroui, and J.R. Gonzalez, *Frailtypack: An r package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation*. Journal of Statistical Software, 2012. **47**.
166. Garcia-Hernandez, A. and D. Rizopoulos, *%JM: A SAS macro to fit jointly generalized mixed models for longitudinal data and time-to-event responses*. Journal of Statistical Software, 2015. **to appear**.
167. Zhang, D., et al., *JMFit: A SAS Macro for Assessing Model Fit in Joint Models of Longitudinal and Survival Data*. Journal of Statistical Software, 2015. **in press**.
168. Lun, D.J., et al., *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility*. STATISTICS AND COMPUTING, 2000. **10**(4): p. 325-337.
169. Beal, S.L., S. L.B., and *NONMEM Users Guides*. Icon Development Solutions, Ellicott City, Maryland, USA. , 1989-98.
170. Smyth, G.K., *Polynomial approximation*. In: Encyclopedia of Biostatistics, P. Armitage and T. Colton (eds.) Wiley, London, 1998: p. 3425 3429.

171. Muthén, L.K. and B.O. Muthén, *Mplus User's Guide: Statistical Analysis with Latent Variables (7th ed.)*, in Los Angeles, CA, M. Muthén., Editor. 1998-2012.
172. Powney, M., et al., *A review of the handling of missing longitudinal outcome data in clinical trials*. TRIALS, 2014. **15**.
173. Powell, C., et al., *MAGNESium Trial In Children (MAGNETIC): a randomised, placebo-controlled trial and economic evaluation of nebulised magnesium sulphate in acute severe asthma in children*. Health Technology Assessment, 2013. **17**(32): p. 1-216 216p.
174. Powell, C., et al., *Magnesium sulphate in acute severe asthma in children (MAGNETIC): a randomised, placebo-controlled trial*. The Lancet Respiratory Medicine, 2013. **1**(4): p. 301-308.
175. Royston, P. and M.K.B. Parmar, *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects*. 2002, JOHN WILEY & SONS LTD: Great Britain. p. 2175.
176. Sudell, M., et al., *Investigation of 2-stage meta-analysis methods for joint longitudinal and time-to-event data through simulation and real data application*. Statistics in Medicine, 2017.
177. Kaplan, E.L. and P. Meier, *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association, 1958(282): p. 457.
178. Akaike, H., *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 1974. **19**(6): p. 716-723.
179. Schwarz, G., *Estimating the Dimension of a Model*. The Annals of Statistics, 1978. **6**(2): p. 461-464.
180. Hua, H., et al., *One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information*. Statistics in Medicine, 2017. **36**(5): p. 772-789.
181. Hickey, G.L., et al., *joinerML: Joint Modelling of Multivariate Longitudinal Data and Time-to-Event Outcomes*. R package version 0.3.0. <https://CRAN.R-project.org/package=joinerML>. 2017.
182. Schwarzer, G., *meta: General Package for Meta-Analysis* 2015.
183. Viechtbauer, W., *Conducting meta-analyses in R with the metafor package*. Journal of Statistical Software, 2010. **36**(3): p. 1-48.
184. Higgins, J., S. Green, and (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
185. Rizopoulos, D., G. Verbeke, and G. Molenberghs, *Multiple-Imputation-Based Residuals and Diagnostic Plots for Joint Models of Longitudinal and Survival Outcomes*. Biometrics, 2010. **66**(1): p. 20-29.
186. Park, K.Y. and P. Qiu, *Model selection and diagnostics for joint modeling of survival and longitudinal data with crossing hazard rate functions*. Statistics in Medicine, 2014.
187. Dobson, A. and R. Henderson, *Diagnostics for Joint Longitudinal and Dropout Time Modeling*. Biometrics, 2003(4): p. 741.
188. Williamson, P.R., et al., *The COMET Handbook: version 1.0*. Trials, 2017. **18**(Suppl 3):280.
189. Burke, D.L., J. Ensor, and R.D. Riley, *Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ*. Statistics in Medicine, 2017. **36**(5): p. 855-875.

190. Debray, T.P.A., et al., *Individual Participant Data Meta-Analysis for a Binary Outcome: One-Stage or Two-Stage?* PLoS ONE, 2013. **8**(4): p. 1-10.
191. Hamza, T.H., T. Stijnen, and H.C. van Houwelingen, *The binomial distribution of meta-analysis was preferred to model within-study variability*. JOURNAL OF CLINICAL EPIDEMIOLOGY, 2008. **61**(1): p. 41-51.
192. Fisher, D.J., et al., *A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners*. JOURNAL OF CLINICAL EPIDEMIOLOGY, 2011. **64**(9): p. 949-967.
193. Stijnen, T., T.H. Hamza, and P. Özdemir, *Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data*. Statistics in Medicine, 2010. **29**(29): p. 3046-3067.
194. Debray, T.P.A., et al., *Get real in individual participant data (IPD) meta-analysis: a review of the methodology*. Research synthesis methods. , 2015. **6**(4): p. 293-309.
195. Therneau, T.M., *A Package for Survival Analysis in S*. 2015.
196. Rizopoulos, D., *Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule*. Computational Statistics & Data Analysis, 2012. **56**(3): p. 491-501.
197. Dempster, A.P., N.M. Laird, and D.B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*. 1977, Royal Statistical Society. p. 1.
198. Wickham, H., *Reshaping Data with the reshape Package*. 2007, 2007. **21**(12): p. 20.
199. Wickham, H., *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.6.3. <https://CRAN.R-project.org/package=tidyr>. 2017.
200. Wickham, H., *ggplot2. [electronic book] : elegant graphics for data analysis*. Use R! 2009: Dordrecht ; Springer, 2009.
201. Oehlert, G.W., *A Note on the Delta Method*. 1992, American Statistical Association. p. 27.
202. Liu, L. and X. Huang, *The use of Gaussian quadrature for estimation in frailty proportional hazards models*. Statistics in Medicine, 2008. **27**(14): p. 2665-2683 19p.
203. Meng, X. and D. van Dyk, *Fast EM-type implementations for mixed effects models*. Journal of The Royal Statistical Society Series B Statistical Methodology, 1998. **60**(3): p. 559-578.
204. Bates, D., et al., *Fitting Linear Mixed-Effects Models Using lme4*. JOURNAL OF STATISTICAL SOFTWARE, 2015. **67**(1): p. 1-48.
205. Stroud, A.H. and D. Secrest, *Gaussian quadrature formulas*. Prentice-Hall series in automatic computation. 1966: Englewood Cliffs (N.J.): Prentice-Hall, 1966.
206. Abramowitz, M. and I.A. Stegun, *Handbook of mathematical functions with formulas, graphs and mathematical tables*. 1964: New York: Dover, 1964.
207. Golub, G.H., *Some Modified Matrix Eigenvalue Problems*. 1973, The Society for Industrial and Applied Mathematics. p. 318.
208. Golub, G.H. and J.H. Welsch, *Calculation of Gauss Quadrature Rules*. 1969, American Mathematical Society. p. 221.
209. Pinheiro, J.C. and D.M. Bates, *Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model*. 1995, American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America. p. 12.
210. Liu, Q.F. and D.A. Pierce, *A Note on Gauss-Hermite Quadrature*. 1994, Biometrika Trust. p. 624.
211. Lesaffre, E. and B. Spiessens, *On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example*. 2001, Blackwell Publishers. p. 325.

212. Rabe-Hesketh, S., A. Skrondal, and A. Pickles, *Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects*. Journal of Econometrics, 2005. **128**: p. 301-323.
213. Pinheiro, J.C. and E.C. Chao, *Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models*. 2006, American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America. p. 58.
214. Rabe-Hesketh, S., A. Skrondal, and A. Pickles, *Reliable estimation of generalized linear mixed models using adaptive quadrature*. The Stata Journal, 2002. **2**(1): p. 1-21.
215. Gibbons, R.D. and D. Hedeker, *Random Effects Probit and Logistic Regression Models for Three-Level Data*. 1997, International Biometric Society. p. 1527.
216. Hedeker, D. and R.D. Gibbons, *A Random-Effects Ordinal Regression Model for Multilevel Analysis*. 1994, International Biometric Society. p. 933.
217. Raman, R. and D. Hedeker, *A mixed-effects regression model for three-level ordinal response data*. STATISTICS IN MEDICINE, 2005. **24**(21): p. 3331-3345.
218. Anderson, T.W. and I. Olkin, *Maximum-Likelihood Estimation of the Parameters of a Multivariate Normal Distribution*. Linear Algebra and its Applications, 1985. **70**: p. 147-171.
219. Bender, R., T. Augustin, and M. Blettner, *Generating survival times to simulate Cox proportional hazards models*. STATISTICS IN MEDICINE, 2005. **24**: p. 1713-1723.
220. Austin, P.C., *Generating survival times to simulate Cox proportional hazards models with time-varying covariates*. STATISTICS IN MEDICINE, 2012. **31**: p. 3946-3958.
221. Amery, A., et al., *Mortality and morbidity results from the European Working Party on High Blood Pressure in the Elderly Trial*. Lancet, 1985. **i**: p. 1349-1354.
222. Coope, J. and T.S. Warrender, *Randomised trial of treatment of hypertension in elderly patients in primary care*. BMJ, 1986. **293**: p. 1145-1151.
223. Dalhöf, B., et al., *Morbidity and mortality in the Swedish Trial in Old Patients with Hypertension (STOP - Hypertension)*. Lancet, 1991. **338**: p. 1281-1285.
224. SHEP Cooperative Research Group, *Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP)*. JAMA, 1991. **265**: p. 3255-3264.
225. Medical Research Council Working Party, *MRC Trial of treatment of mild hypertension: principal results*. BMJ, 1985. **291**: p. 97-104.
226. MRC Working Party, *Medical Research Council trial of treatment of hypertension in older adults: principal results*. BMJ, 1992. **304**: p. 405-412.
227. *Concise medical dictionary*. 9 ed. Oxford Reference Online., ed. E.A. Martin. 2015: Oxford University Press.
228. National Institute for Health and Care Excellence. *Cardiovascular disease prevention NICE guideline. (PH25)*.
229. National Institute for Health and Care Excellence. *Hypertension in adults: diagnosis and management NICE guideline. (CG127)*.
230. Collins, G.S. and D. Altman, G., *An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study*. BMJ: British Medical Journal., 2009. **339**(7713): p. 144-147.
231. Collins, G.S. and D.G. Altman, *An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study*. BMJ: British Medical Journal., 2010. **340**(2442).

232. Collins, G.S. and D.G. Altman, *Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2.* BMJ (Clinical Research Ed.), 2012. **344**(4181).
233. Hippisley-Cox, J., et al., *Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: A validation study.* Heart, 2008. **94**(1): p. 34-39.
234. Hippisley-Cox, J., et al., *Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study.* BMJ: British Medical Journal., 2007. **335**(7611): p. 136-141.
235. Hippisley-Cox, J., et al., *Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2.* BMJ: British Medical Journal. , 2008. **336**(7659): p. 1475-1482.
236. Reckelhoff, J.F., *Gender Differences in the Regulation of Blood Pressure.* Hypertension, 2001. **37**(5): p. 1199-1208.
237. Reule, S. and P.E. Drawz, *Heart Rate and Blood Pressure: Any Possible Implications for Management of Hypertension?* Current hypertension reports, 2012. **14**(6): p. 478-484.
238. Schroeder, E.B., et al., *Hypertension, Blood Pressure, and Heart Rate Variability: The Atherosclerosis Risk in Communities (ARIC) Study.* Hypertension, 2003. **42**: p. 1106-1111.
239. National Institute for Health and Care Excellence. *Chronic kidney disease in adults: assessment and management. NICE guideline (CG182).*
240. Litzkow, M.J., M. Livny, and M.W. Mutka. *Condor - a hunter of idle workstations.* 1988. Publ by IEEE.
241. Bretagnolle, J. and C. Huber-Carol, *Effects of Omitting Covariates in Cox's Model for Survival Data.* Scandinavian Journal of Statistics 1988. **15**(2): p. 125-138.
242. Abrahamowicz, M., et al., *Bias due to Aggregation of Individual Covariates in the Cox Regression Model.* American Journal of Epidemiology, 2004. **160**(7): p. 696-706.
243. SCHMOOR, C.a.S., M., *EFFECTS OF COVARIATE OMISSION AND CATEGORIZATION WHEN ANALYSING RANDOMIZED TRIALS WITH THE COX MODEL.* STATISTICS IN MEDICINE, 1997. **16**: p. 225-237.
244. Sweeting, M.J. and S.G. Thompson, *Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture.* BIOMETRICAL JOURNAL, 2011. **53**(5): p. 750-763.
245. Lin, H.Q., C.E. McCulloch, and S.T. Mayne, *Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables.* STATISTICS IN MEDICINE, 2002. **21**(16): p. 2369-2382.
246. Williamson, P.R., et al., *Joint modelling of longitudinal and competing risks data.* STATISTICS IN MEDICINE, 2008. **27**(30): p. 6426-6438.
247. Nevitt, S.J., et al., *Antiepileptic drug monotherapy for epilepsy: a network meta-analysis of individual participant data.* The Cochrane Database Of Systematic Reviews, 2017. **6**: p. CD011412.
248. Rizopoulos, D., et al., *Combining Dynamic Predictions from Joint Models for Longitudinal and Time-to-Event Data using Bayesian Model Averaging.* 2013.
249. Rizopoulos, D., et al., *Dynamic Predictions with Time-Dependent Covariates in Survival Analysis using Joint Modeling and Landmarking.* 2013.
250. Rizopoulos, D., G. Molenberghs, and E.M.E.H. Lesaffre, *Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking.* Biometrical Journal. Biometrische Zeitschrift, 2017. **59**(6): p. 1261-1276.



251. Vermunt, J.K., *An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models*. Statistica Neerlandica, 2004. **58**(2): p. 220-233.

# Appendices

---

## Appendix 1: Search Strategies used in Review of Current Reporting Standards of Joint Models in the Literature

Database : Scopus

```
( TITLE-ABS-KEY ( joint W/3 model* ) ) AND ( ( TITLE-ABS-KEY ( longitudinal W/4 survival ) )  
OR ( TITLE-ABS-KEY ( longitudinal W/4 "time-to-event" ) ) OR ( TITLE-ABS-KEY ( longitudinal  
W/4 ( time W/3 event ) ) ) OR ( TITLE-ABS-KEY ( "repeat* measure*" W/4 survival ) ) OR ( TITLE-ABS-KEY ( "repeat* measure*" W/4 "time-to-event" ) ) OR ( TITLE-ABS-KEY ( "repeat* measure*" W/4 ( time W/3 event ) ) ) )
```

Database : PubMed

```
(joint model*) AND (((((((longitudinal and survival)) OR (longitudinal and "time-to-event"))  
OR (longitudinal and "time to event")) OR (longitudinal and "event time")) OR ((repeat*  
measure*) and survival)) OR ((repeat* measure*) and "time-to-event")) OR ((repeat*  
measure*) and "time to event")) OR ((repeat* measure*) and "event time"))
```

Database: Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) <1946 to Present with Daily Update>

Search Strategy:

- 
- 1 (joint adj3 model\*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2329)
  - 2 (longitudinal adj4 survival).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (283)

3 (longitudinal adj4 time-to-event).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (39)

4 (longitudinal adj4 (time adj3 event)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (54)

5 ("repeat\* measure\*" adj4 survival).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (37)

6 ("repeat\* measure\*" adj4 time-to-event).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (10)

7 ("repeat\* measure\*" adj4 (time adj3 event)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (15)

8 2 or 3 or 4 or 5 or 6 or 7 (367)

9 1 and 8 (102)

\*\*\*\*\*

## Appendix 2: Additional longitudinal SBP trajectory plots for INDANA dataset

SBP and time to Death

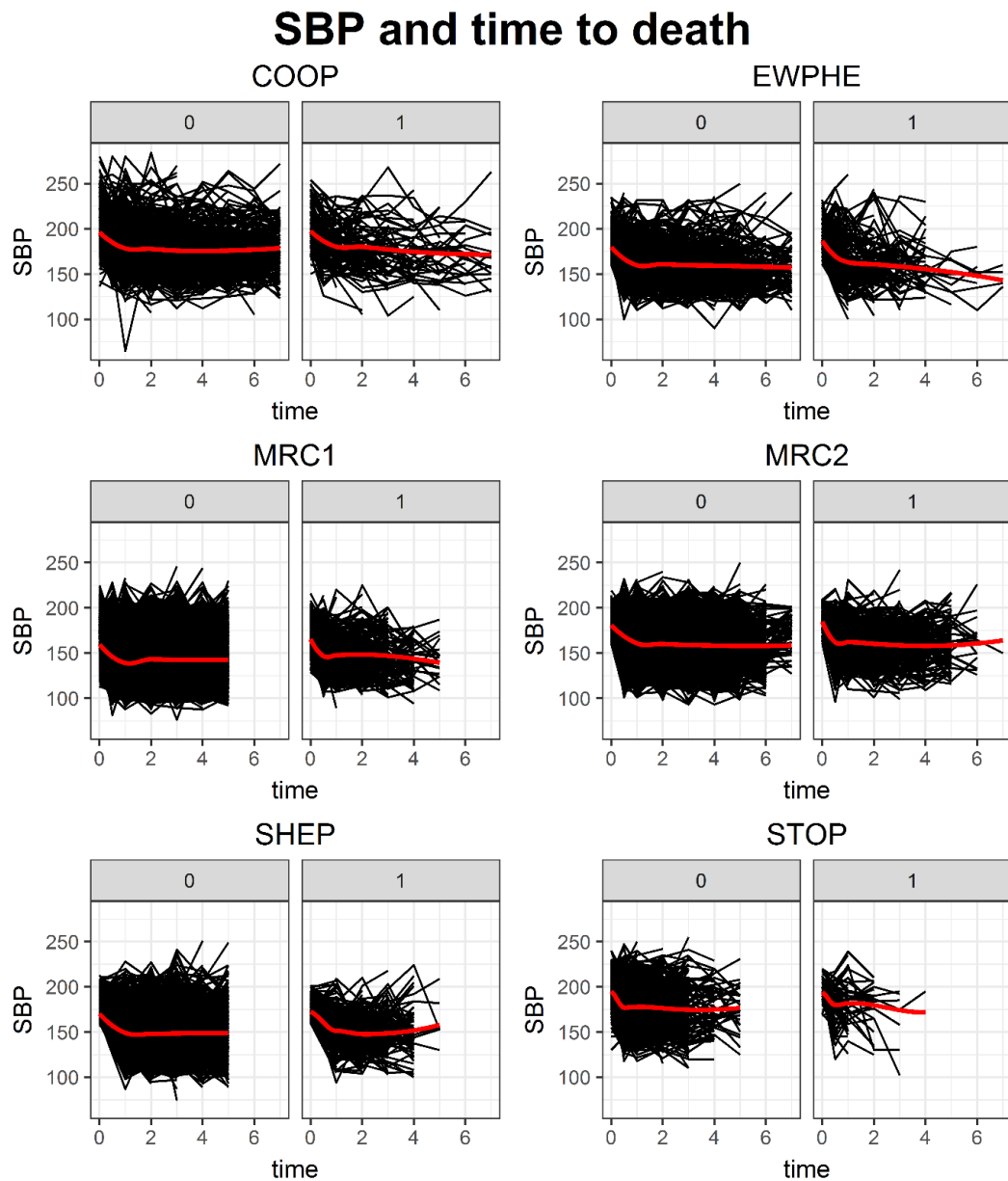


Figure 37: Longitudinal trajectory plots with mean trajectory smoother (red line) for SBP and time to death data, with 0 indicating a censoring and 1 indicating an event was experienced.

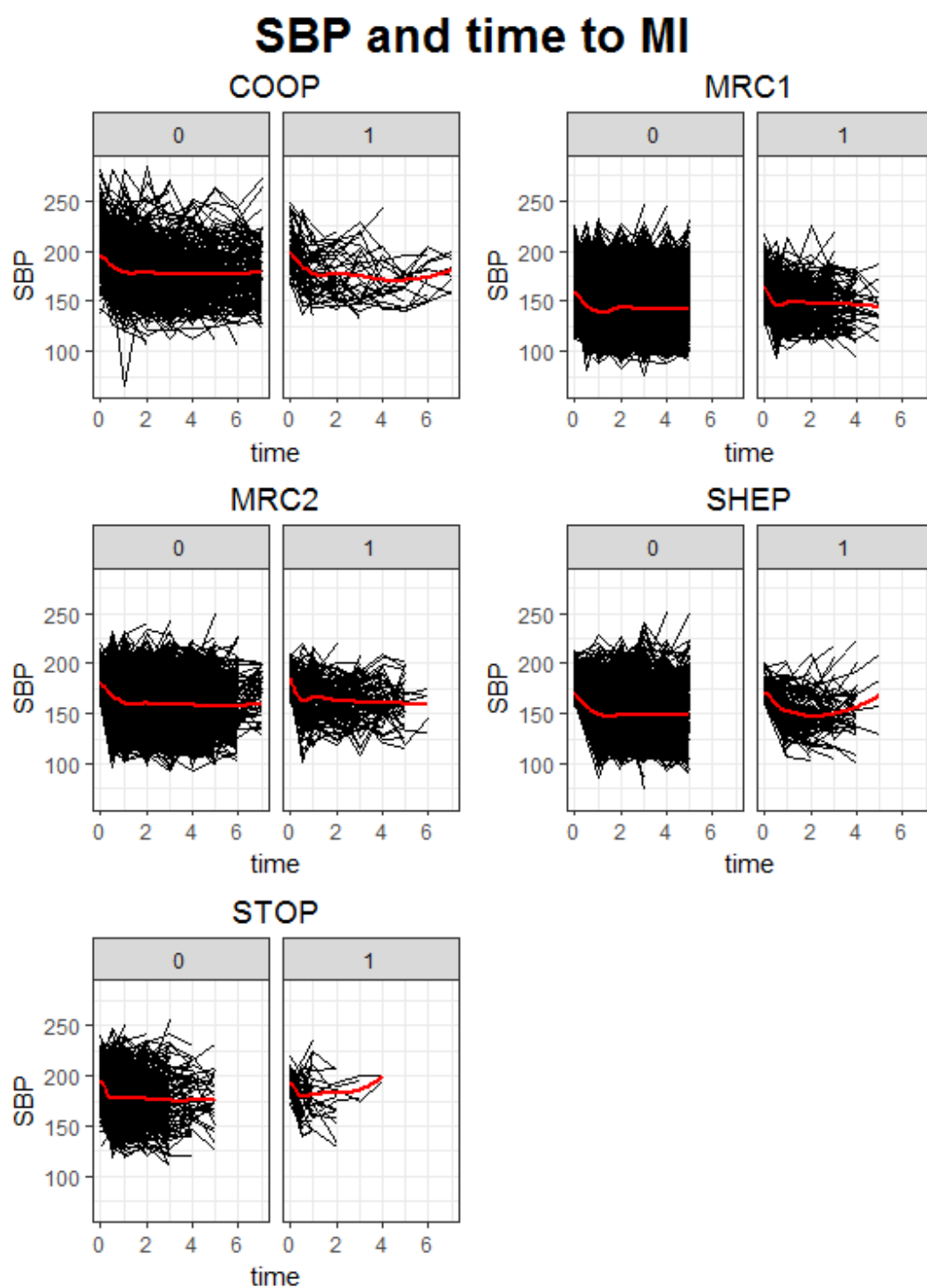


Figure 38: Longitudinal trajectory plots with mean trajectory smoother (red line) for SBP and time to myocardial infarction (MI) data, with 0 indicating a censoring and 1 indicating an event was experienced.

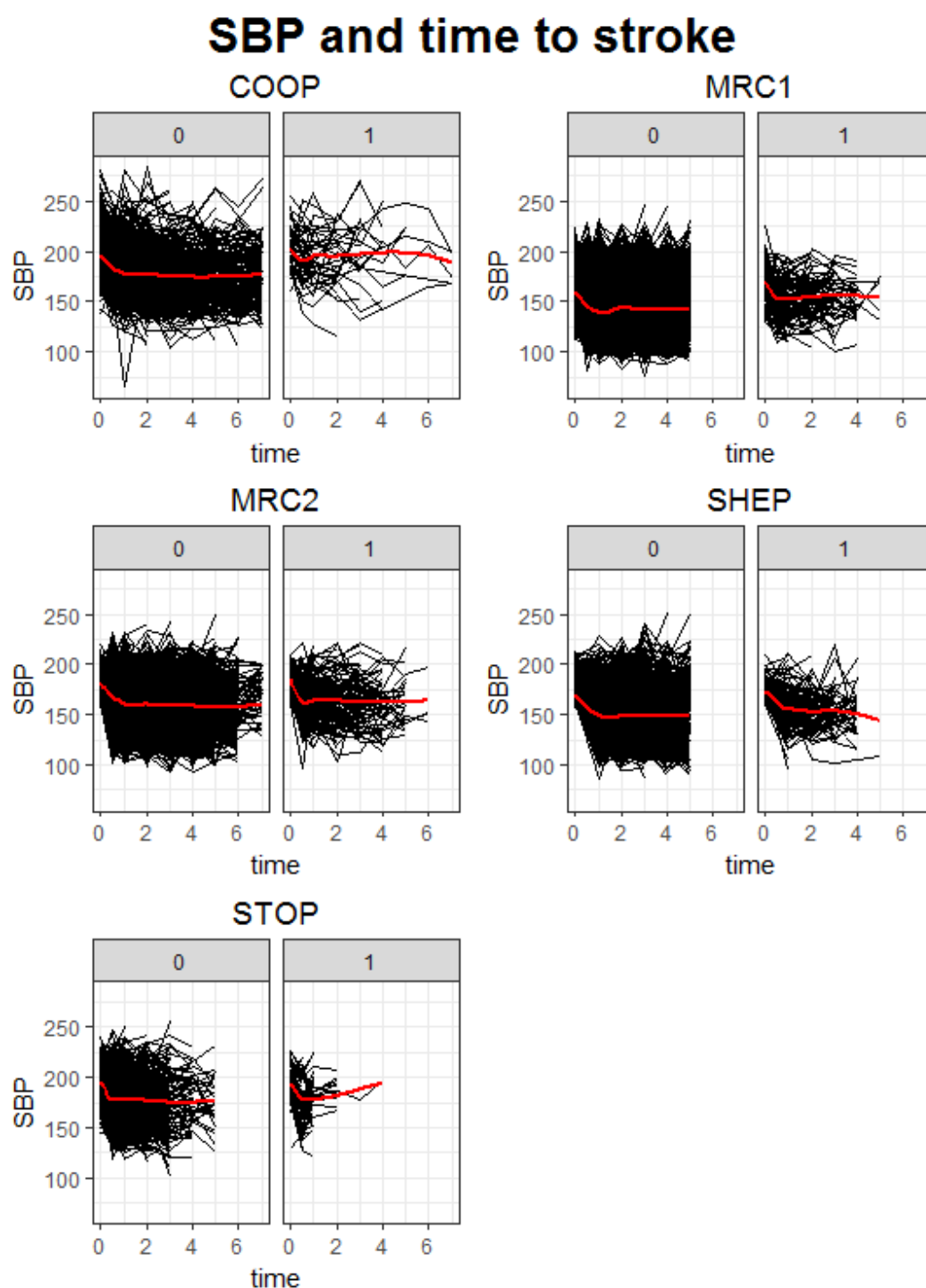


Figure 39: Longitudinal trajectory plots with mean trajectory smoother (red line) for SBP and time to stroke data, with 0 indicating a censoring and 1 indicating an event was experienced.

### Appendix 3: Event Rates for Two-Stage Simulation Studies

Scenario		Measures	Overall	Study 1	Study 2	Study 3	Study 4	Study 5
	$\alpha^{(2)}$							
Low Event rate Homogeneous	0	Mean (sd)	24 (0.8)	24 (1.9)	24 (1.9)	24.1 (1.9)	24 (1.8)	24 (1.9)
		Median (LQ,UQ)	24 (23.4,24.6)	24 (22.8,25.2)	24 (22.8,25.2)	24.2 (22.9,25.6)	24 (22.8,25.2)	24 (22.8,25.4)
	0.25	Mean (sd)	23.9 (0.8)	23.9 (1.8)	23.9 (1.9)	24 (1.9)	23.9 (1.9)	23.9 (1.9)
		Median (LQ,UQ)	23.9 (23.4,24.5)	23.8 (22.8,25.2)	24 (22.6,25.2)	24 (22.6,25.2)	23.8 (22.6,25.2)	24 (22.8,25.2)
	0.5	Mean (sd)	24 (0.9)	24 (1.9)	23.9 (1.9)	23.9 (1.9)	24 (1.9)	24 (2)
		Median (LQ,UQ)	24 (23.4,24.5)	24 (22.8,25.4)	23.8 (22.6,25.2)	24 (22.8,25.2)	23.8 (22.8,25.2)	24 (22.6,25.4)
	0.75	Mean (sd)	24.3 (0.9)	24.3 (2)	24.3 (1.8)	24.4 (1.9)	24.4 (2)	24.2 (1.9)
		Median (LQ,UQ)	24.3 (23.7,25)	24.4 (23,25.6)	24.2 (23.2,25.6)	24.3 (23,25.8)	24.4 (23,25.8)	24.2 (23,25.4)
	1	Mean (sd)	24 (0.9)	24 (1.9)	23.9 (1.9)	23.9 (1.9)	24 (1.9)	24 (2)
		Median (LQ,UQ)	24 (23.4,24.5)	24 (22.8,25.4)	23.8 (22.6,25.2)	24 (22.8,25.2)	23.8 (22.8,25.2)	24 (22.6,25.4)
High Event Rate Homogeneous	0	Mean (sd)	87 (0.7)	87 (1.6)	87 (1.5)	87 (1.5)	86.9 (1.6)	87 (1.5)
		Median (LQ,UQ)	87 (86.5,87.4)	87 (86,88)	87 (86,88)	87 (86,88.2)	87 (85.8,88)	87 (86,88)
	0.25	Mean (sd)	86.6 (0.7)	86.6 (1.5)	86.6 (1.5)	86.6 (1.5)	86.6 (1.5)	86.5 (1.5)
		Median (LQ,UQ)	86.6 (86.1,87)	86.6 (85.6,87.6)	86.6 (85.6,87.6)	86.6 (85.6,87.6)	86.6 (85.6,87.6)	86.6 (85.6,87.6)
	0.5	Mean (sd)	84.3 (0.7)	84.3 (1.7)	84.4 (1.7)	84.3 (1.5)	84.3 (1.6)	84.3 (1.7)
		Median (LQ,UQ)	84.3 (83.8,84.8)	84.4 (83.2,85.4)	84.4 (83.2,85.4)	84.2 (83.2,85.4)	84.4 (83.2,85.4)	84.4 (83.2,85.6)
	0.75	Mean (sd)	79.6 (0.8)	79.6 (1.8)	79.6 (1.8)	79.6 (1.8)	79.6 (1.8)	79.7 (1.8)
		Median (LQ,UQ)	79.7 (79.1,80.2)	79.6 (78.4,80.8)	79.6 (78.4,81)	79.6 (78.4,80.8)	79.6 (78.4,80.8)	79.8 (78.6,81)
	1	Mean (sd)	75 (0.9)	75 (1.9)	75 (2)	74.9 (2.1)	75.1 (1.9)	74.9 (1.9)
		Median (LQ,UQ)	75 (74.4,75.5)	75 (73.6,76.2)	75 (73.6,76.2)	74.8 (73.6,76.4)	75.2 (73.8,76.4)	74.8 (73.6,76.2)
Low Event rate Heterogeneous	0	Mean (sd)	24.1 (1)	24.1 (2.1)	24.1 (2.2)	24.2 (2.1)	24.2 (2.1)	24.1 (2.1)
		Median (LQ,UQ)	24.1 (23.5,24.8)	24.1 (22.6,25.6)	24.2 (22.6,25.6)	24.2 (22.8,25.6)	24.2 (22.8,25.6)	24 (22.6,25.6)
	0.25	Mean (sd)	24 (0.9)	23.9 (2.2)	24 (2.2)	24 (2.1)	24 (2.2)	24.1 (2.1)
		Median (LQ,UQ)	24 (23.4,24.6)	23.8 (22.4,25.4)	24 (22.6,25.6)	24 (22.6,25.4)	24 (22.6,25.4)	24 (22.6,25.6)
	0.5	Mean (sd)	24.1 (0.9)	24.1 (2.2)	24.1 (2)	24.2 (2.1)	24.1 (2.2)	24.1 (2.1)
		Median (LQ,UQ)	24.1 (23.5,24.8)	24 (22.6,25.6)	24 (22.6,25.6)	24.2 (22.8,25.6)	24 (22.6,25.6)	24 (22.8,25.6)
	0.75	Mean (sd)	24.4 (0.9)	24.5 (2.2)	24.4 (2.1)	24.4 (2)	24.4 (2.1)	24.3 (2.2)
		Median (LQ,UQ)	24.4 (23.8,25)	24.6 (23.1,26)	24.4 (23,25.8)	24.4 (23,25.8)	24.4 (23,25.8)	24.2 (22.8,25.8)
	1	Mean (sd)	25 (0.9)	25 (2.1)	25.1 (2.1)	25.1 (2.2)	25.2 (2.1)	24.9 (2.1)
		Median (LQ,UQ)	25 (24.4,25.7)	25 (23.8,26.4)	25 (23.6,26.4)	25 (23.6,26.6)	25 (23.8,26.4)	24.9 (23.6,26.2)
High Event Rate Heterogeneous	0	Mean (sd)	87 (0.7)	87 (1.5)	87 (1.5)	87 (1.5)	87 (1.4)	87 (1.5)
		Median (LQ,UQ)	87 (86.6,87.5)	87 (86,88)	87 (86,88)	87 (86,88.2)	87 (86.2,88)	87 (86,88)
	0.25	Mean (sd)	86.6 (0.7)	86.6 (1.5)	86.6 (1.5)	86.6 (1.5)	86.6 (1.5)	86.6 (1.5)
		Median (LQ,UQ)	86.6 (86.1,87)	86.6 (85.6,87.6)	86.6 (85.6,87.6)	86.6 (85.6,87.6)	86.6 (85.6,87.8)	86.6 (85.6,87.6)
	0.5	Mean (sd)	84.3 (0.7)	84.3 (1.6)	84.4 (1.6)	84.5 (1.6)	84.3 (1.6)	84.3 (1.5)
		Median (LQ,UQ)	84.4 (83.9,84.8)	84.2 (83.4,85.4)	84.4 (83.4,85.4)	84.4 (83.4,85.6)	84.2 (83.2,85.6)	84.4 (83.2,85.4)
	0.75	Mean (sd)	79.7 (0.8)	79.7 (1.8)	79.6 (1.8)	79.6 (1.8)	79.6 (1.8)	79.8 (1.7)
		Median (LQ,UQ)	79.7 (79.2,80.2)	79.6 (78.6,80.8)	79.6 (78.4,80.8)	79.6 (78.4,80.8)	79.6 (78.4,80.8)	79.8 (78.6,81)
	1	Mean (sd)	75 (0.9)	74.9 (2)	75 (1.9)	75 (1.9)	74.9 (2)	75.1 (1.9)
		Median (LQ,UQ)	74.9 (74.4,75.6)	75 (73.6,76.2)	75 (73.6,76.2)	75 (73.8,76.4)	74.8 (73.4,76.2)	75.2 (73.8,76.4)

Table 28: Mean and median event rates for data simulated for two stage analysis grouped by target event rate, whether treatment effect was simulated as homogenous or heterogeneous across studies, and by level of association (given by true association parameter  $\alpha^{(2)}$ )

## Appendix 4: One stage simulation graphs showing point estimates and confidence intervals

### Simulation Set 1: Varying levels of association

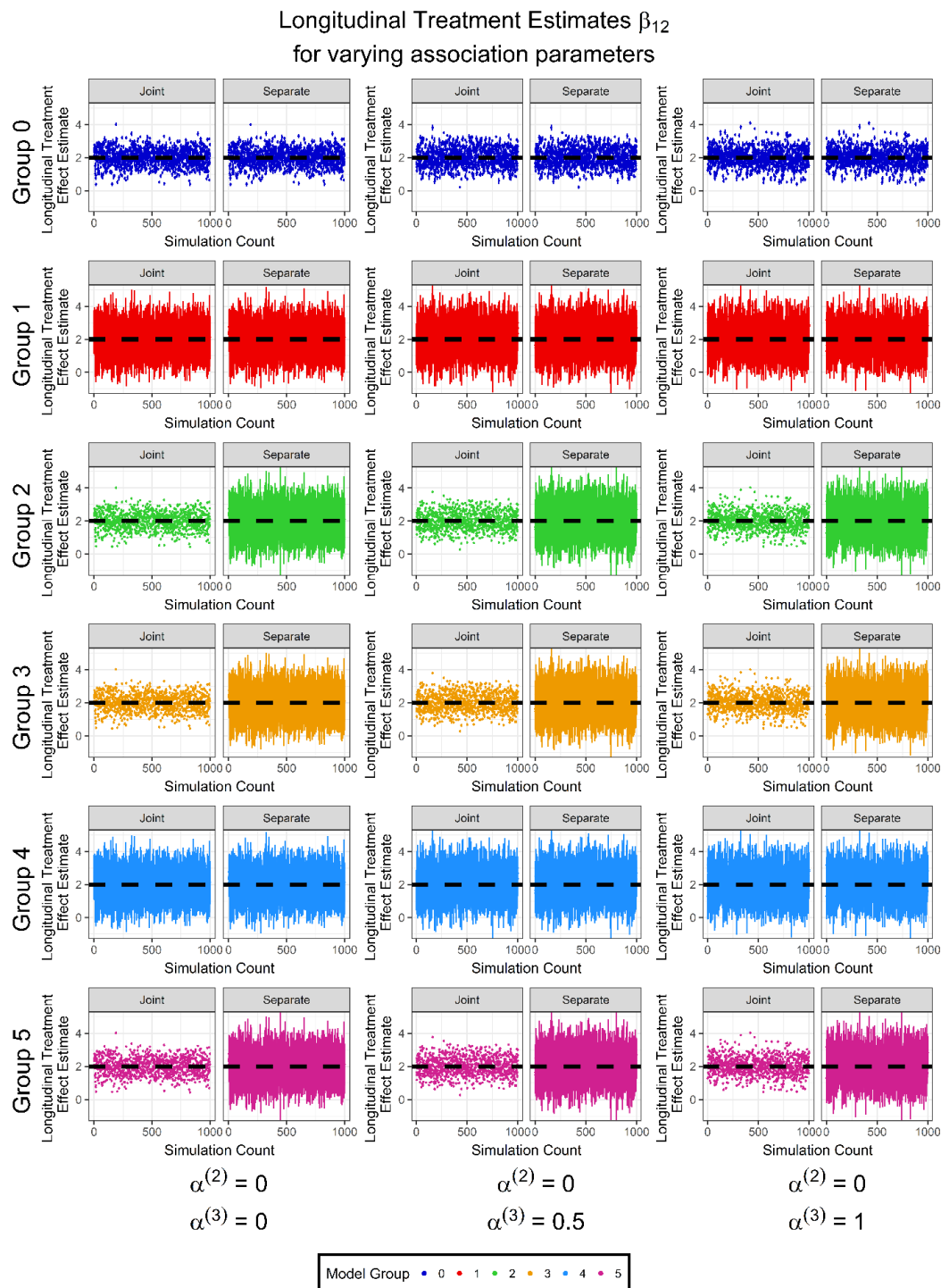


Figure 40: Point estimates and confidence intervals for longitudinal treatment effect parameter ( $\beta_{12}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{12}$  that the data was simulated under.



Longitudinal Treatment Estimates  $\beta_{12}$   
for varying association parameters

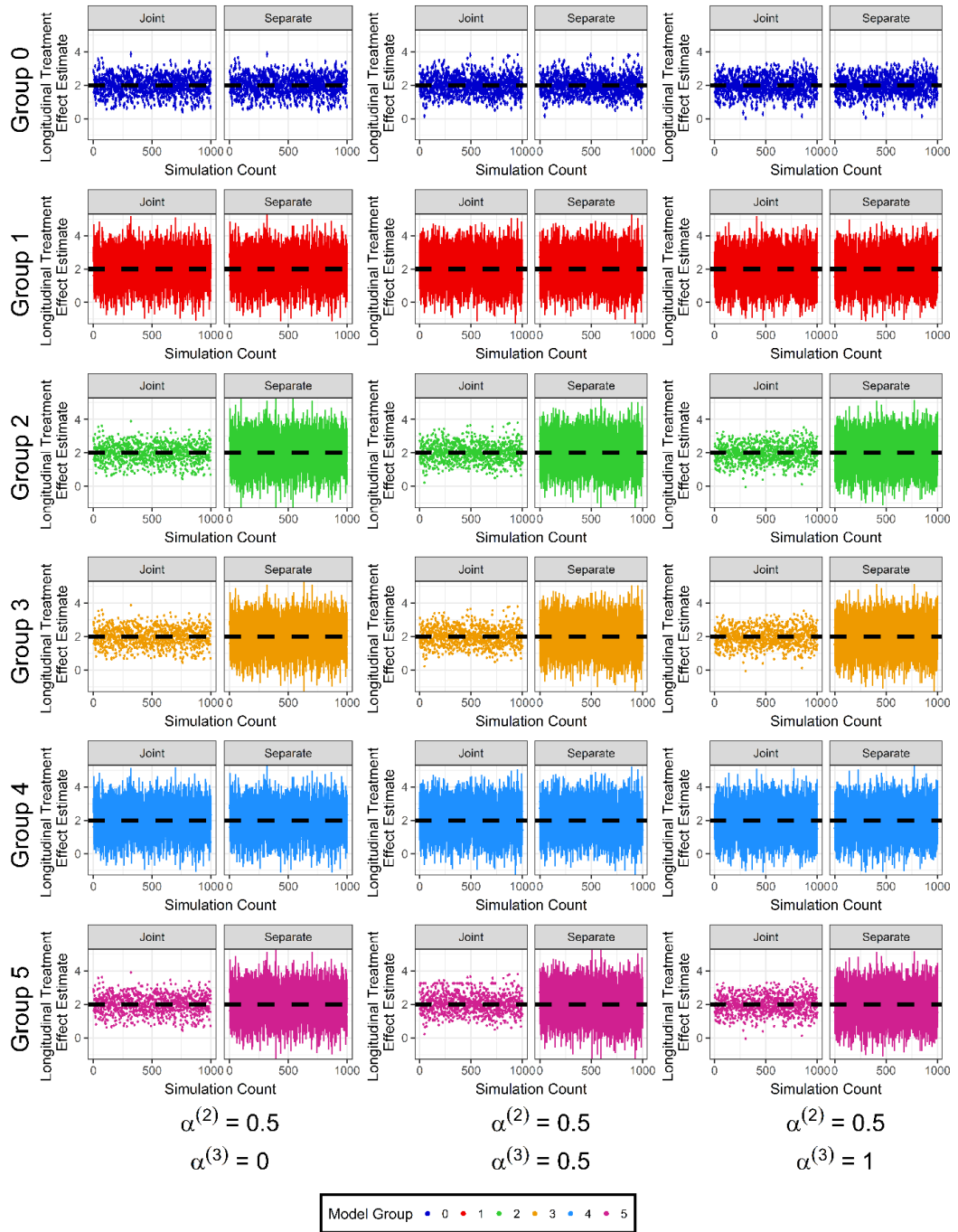


Figure 41: Point estimates and confidence intervals for longitudinal treatment effect parameter ( $\beta_{12}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{12}$  that the data was simulated under.

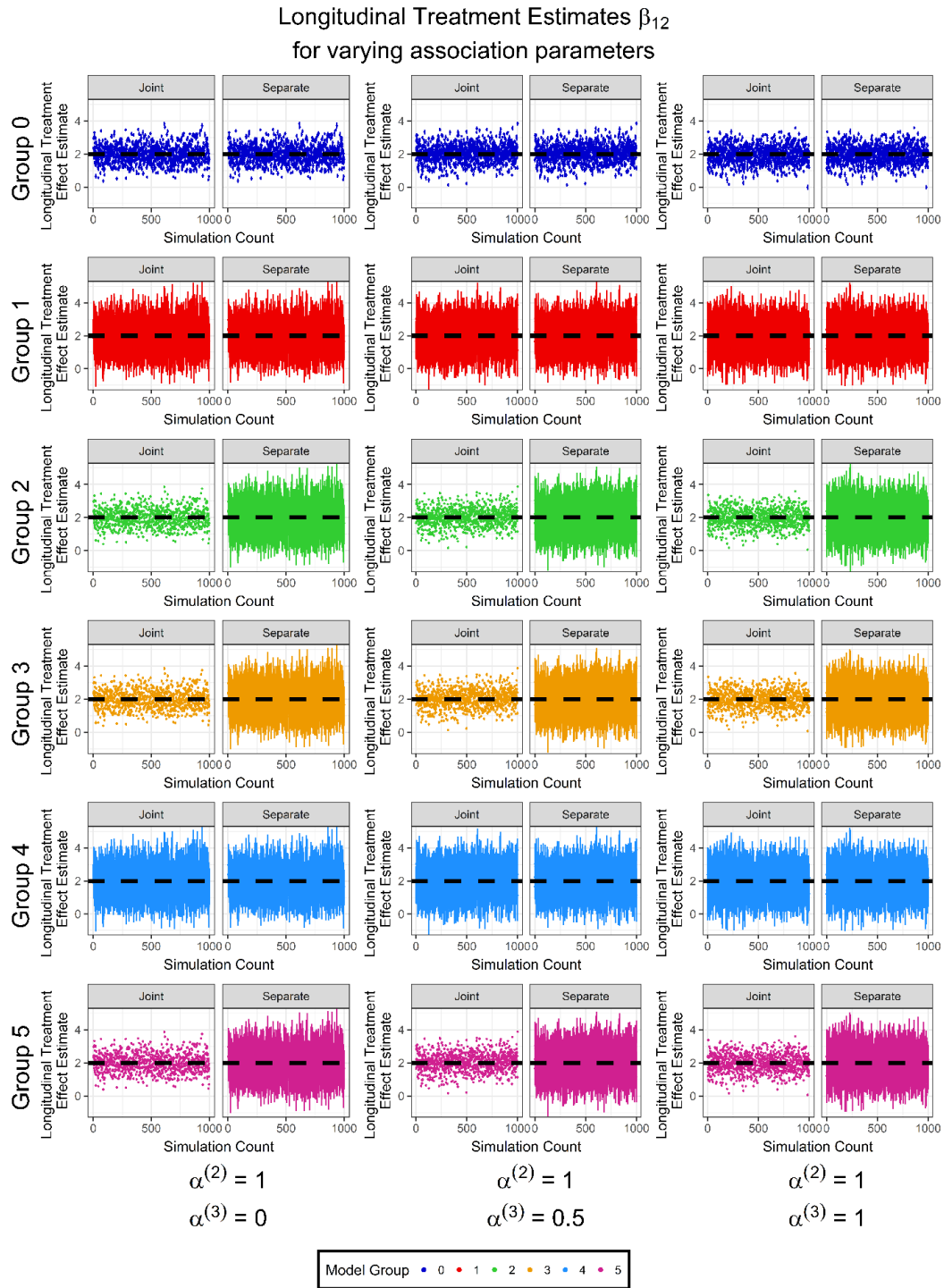


Figure 42: Point estimates and confidence intervals for longitudinal treatment effect parameter ( $\beta_{12}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{12}$  that the data was simulated under.

# Time-to-Event Treatment Estimates $\beta_{21}$ for varying association parameters

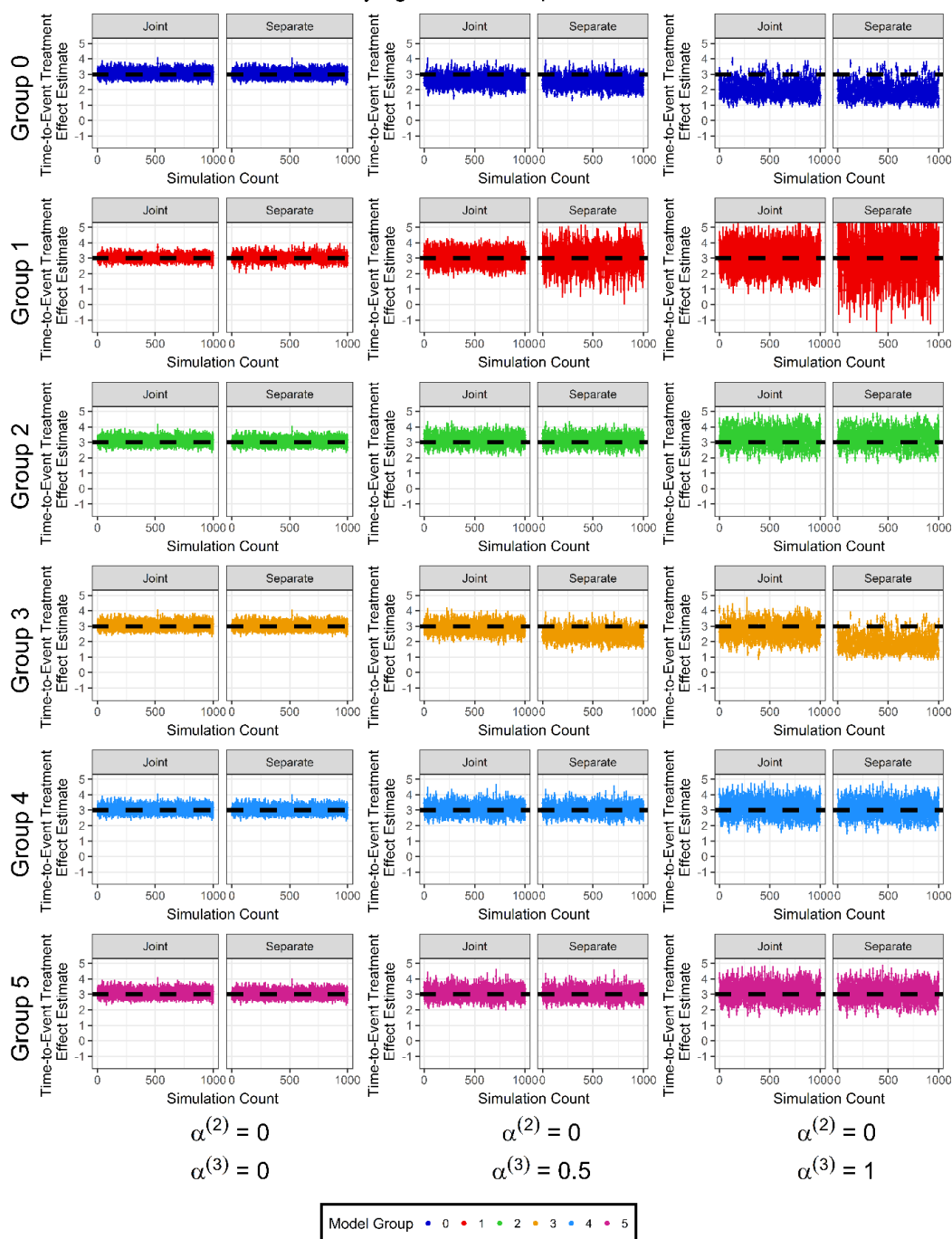


Figure 43: Point estimates and confidence intervals for time-to-event treatment effect parameter ( $\beta_{21}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{21}$  that the data was simulated under.

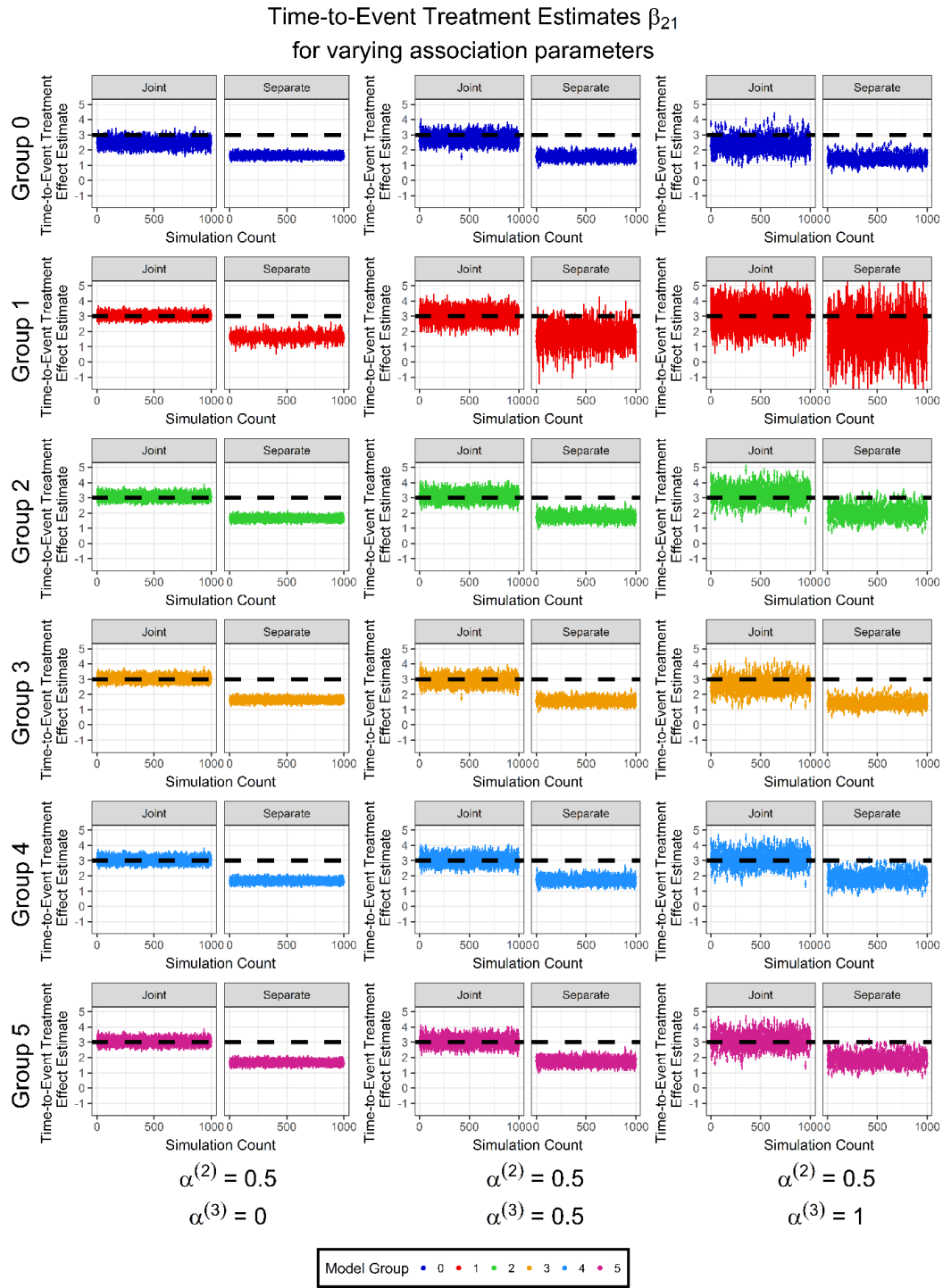


Figure 44: Point estimates and confidence intervals for time-to-event treatment effect parameter ( $\beta_{21}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{21}$  that the data was simulated under.

# Time-to-Event Treatment Estimates $\beta_{21}$ for varying association parameters

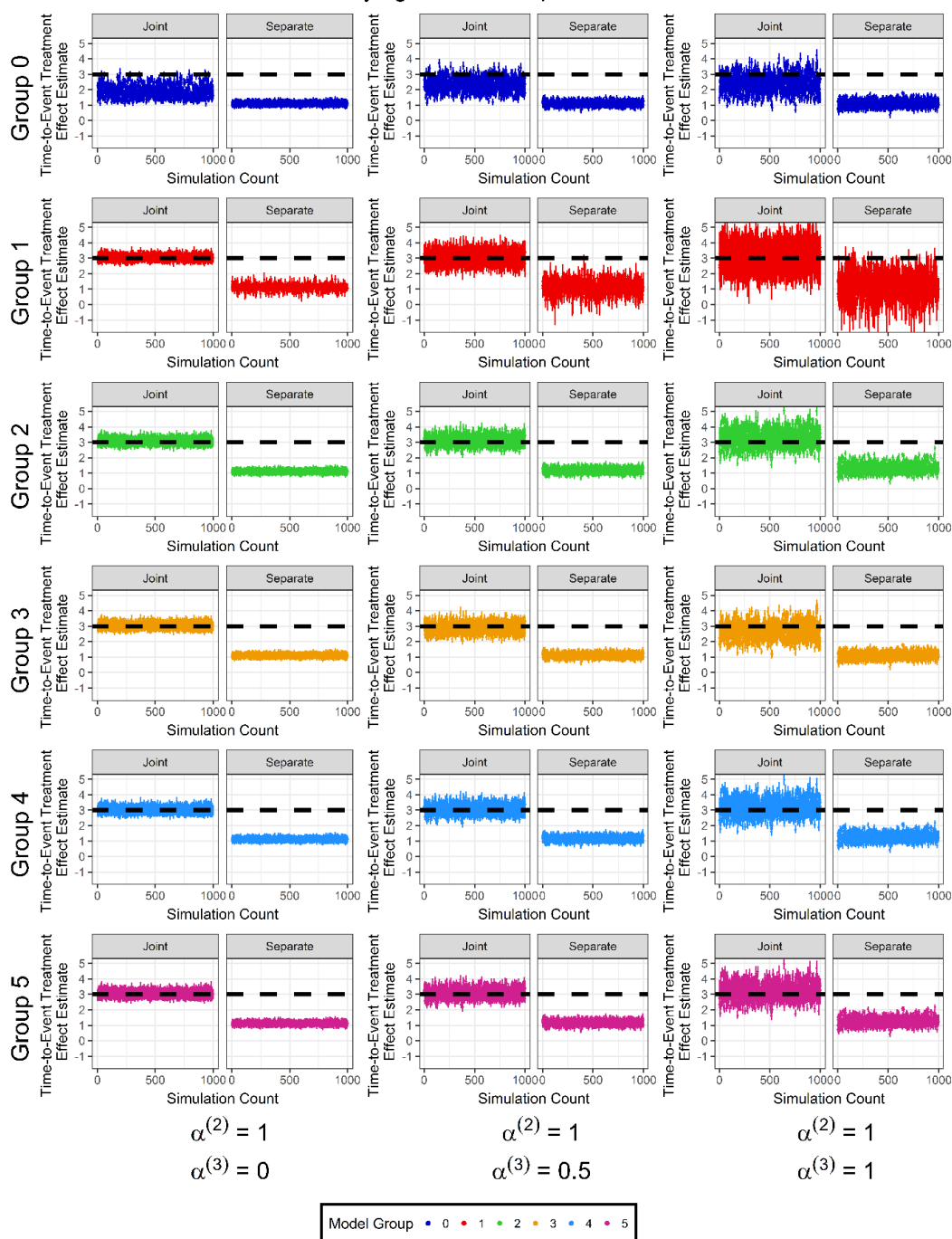


Figure 45: Point estimates and confidence intervals for time-to-event treatment effect parameter ( $\beta_{21}$ ) for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\beta_{21}$  that the data was simulated under.

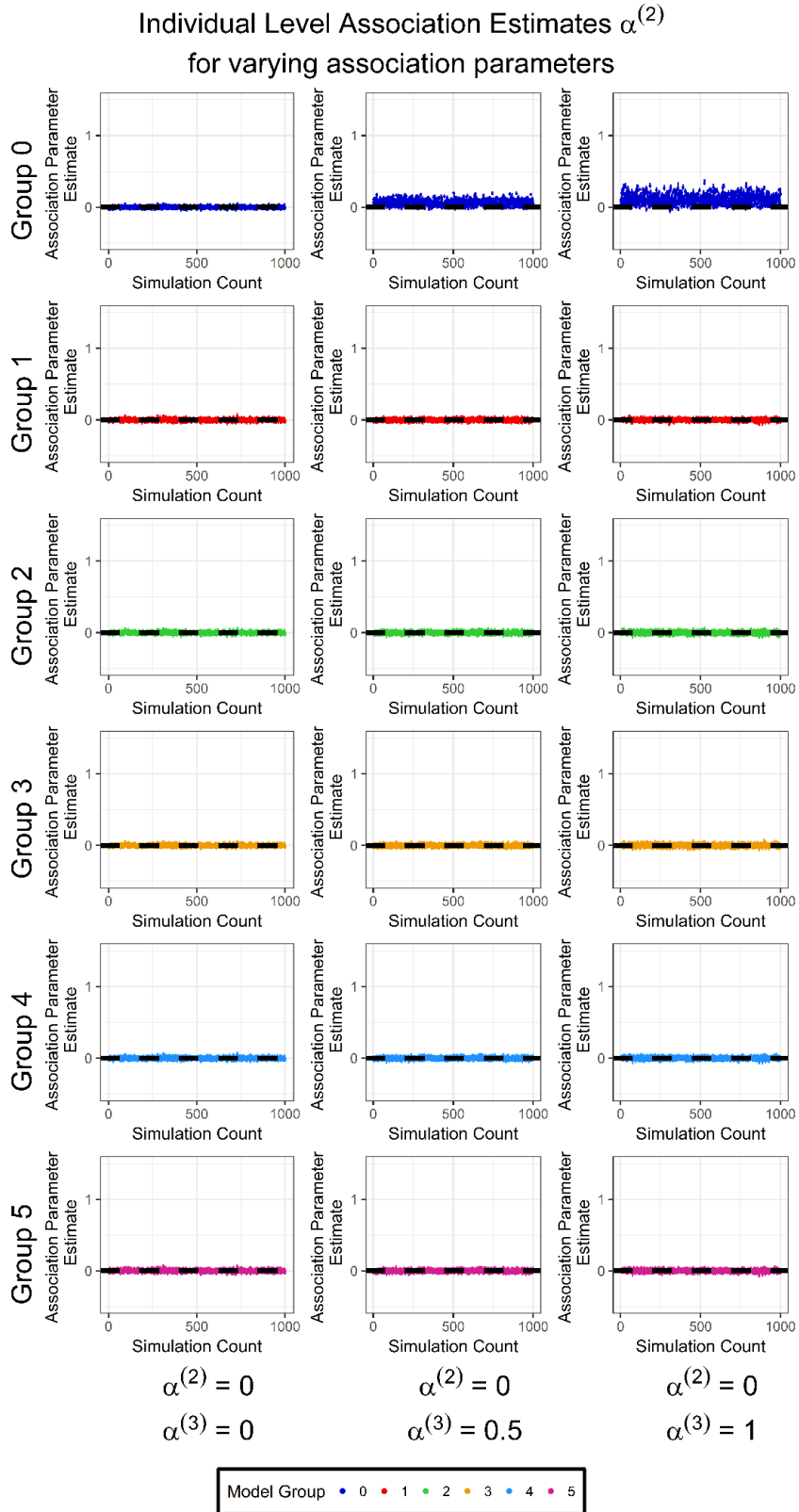


Figure 46: Point estimates and confidence intervals for individual level association parameter  $\alpha^{(2)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(2)}$  that the data was simulated under.

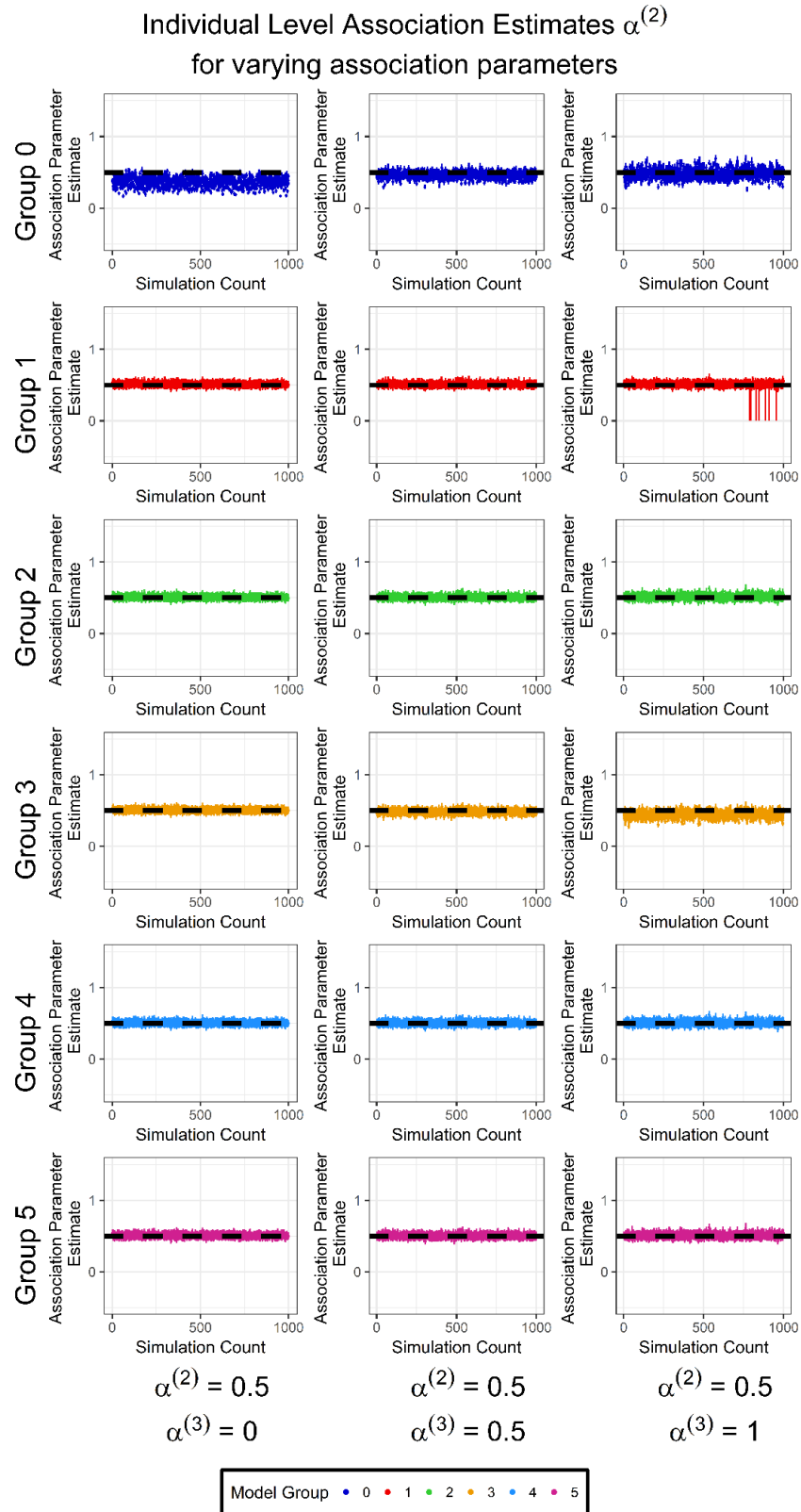


Figure 47: Point estimates and confidence intervals for individual level association parameter  $\alpha^{(2)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(2)}$  that the data was simulated under.



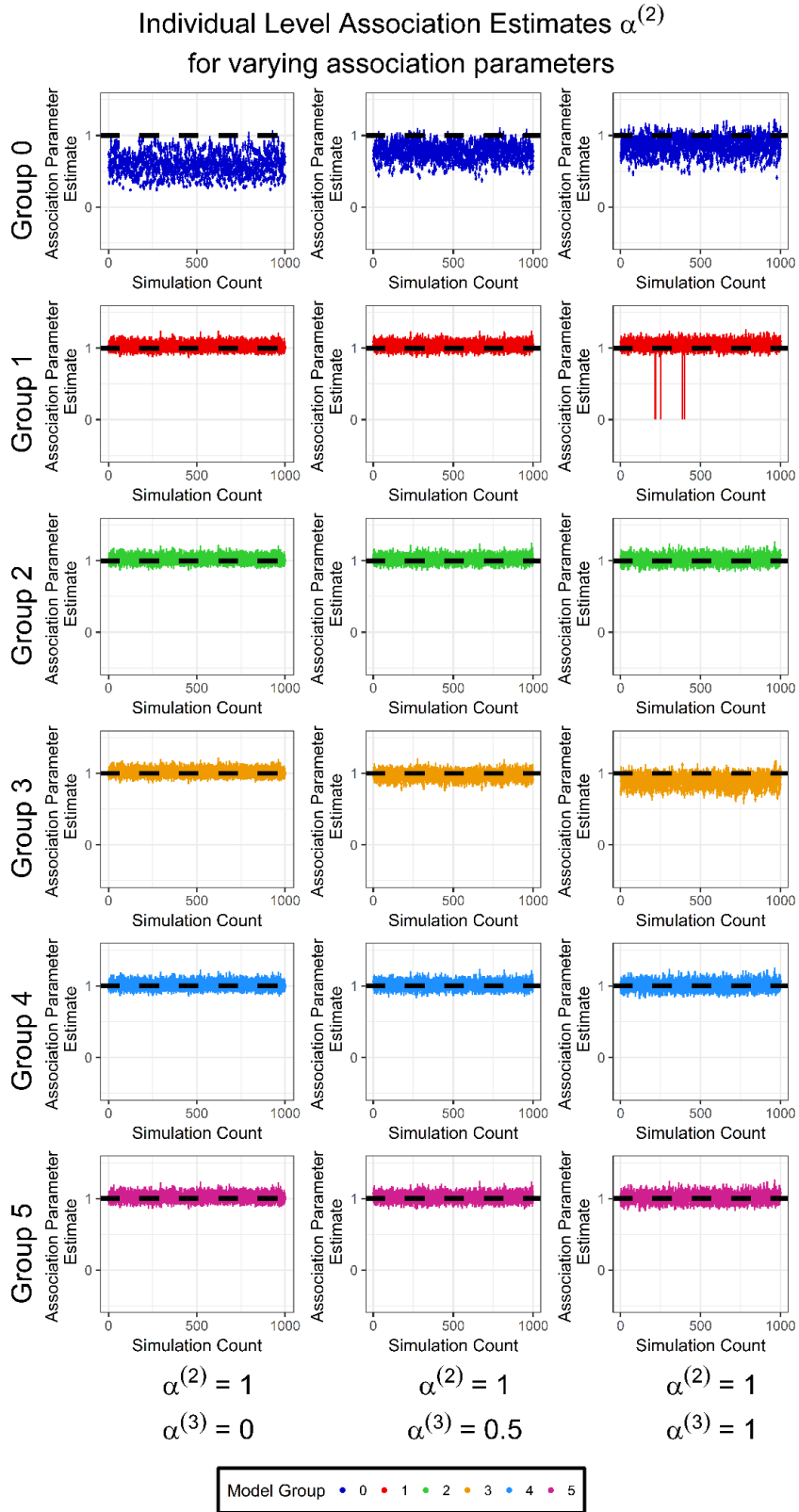


Figure 48: Point estimates and confidence intervals for individual level association parameter  $\alpha^{(2)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(2)}$  that the data was simulated under.



## Study Level Association Estimates $\alpha^{(3)}$ for varying association parameters

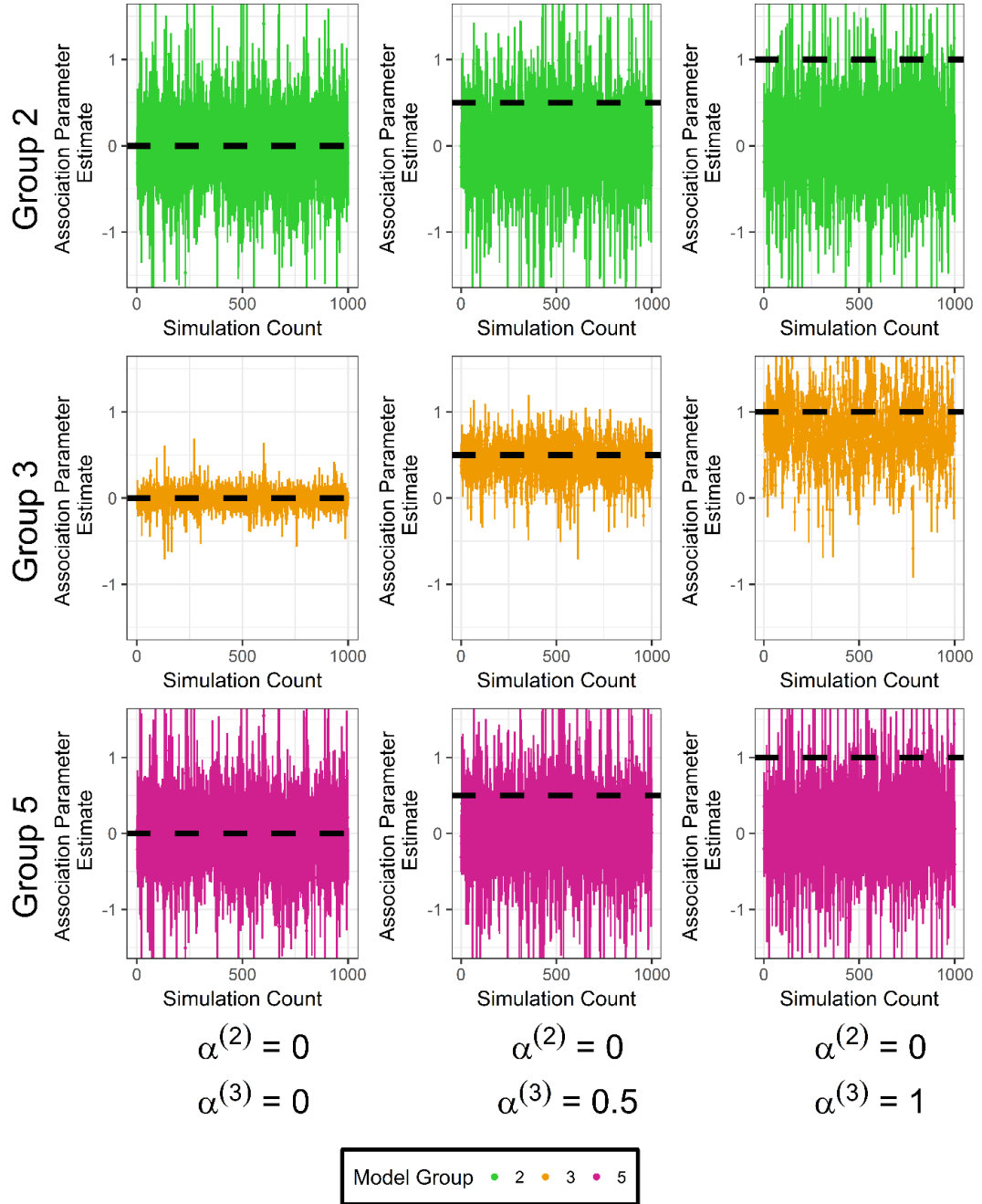


Figure 49: Point estimates and confidence intervals for study level association parameter  $\alpha^{(3)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(3)}$  that the data was simulated under.

## Study Level Association Estimates $\alpha^{(3)}$ for varying association parameters

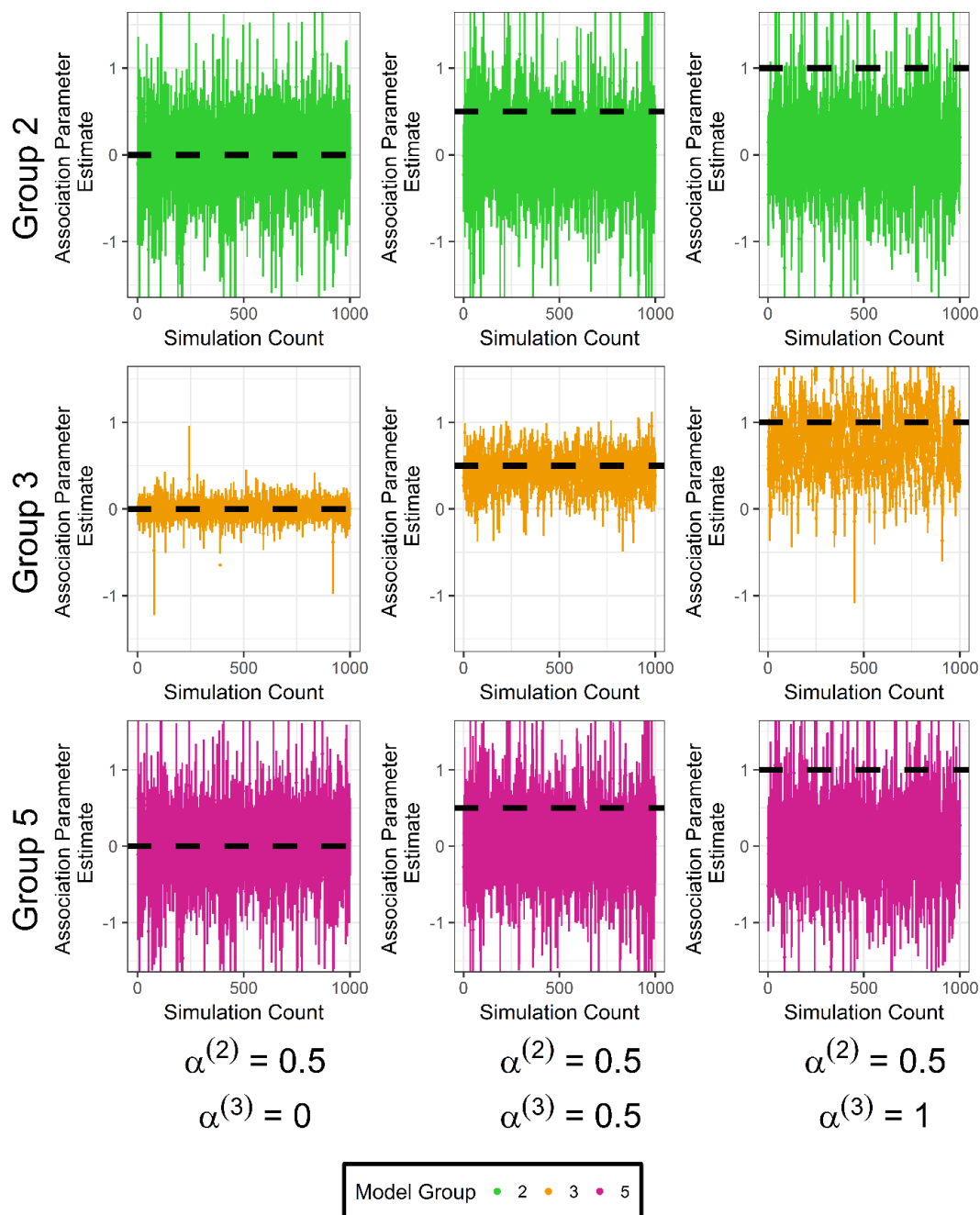


Figure 50: Point estimates and confidence intervals for study level association parameter  $\alpha^{(3)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(3)}$  that the data was simulated under

## Study Level Association Estimates $\alpha^{(3)}$ for varying association parameters

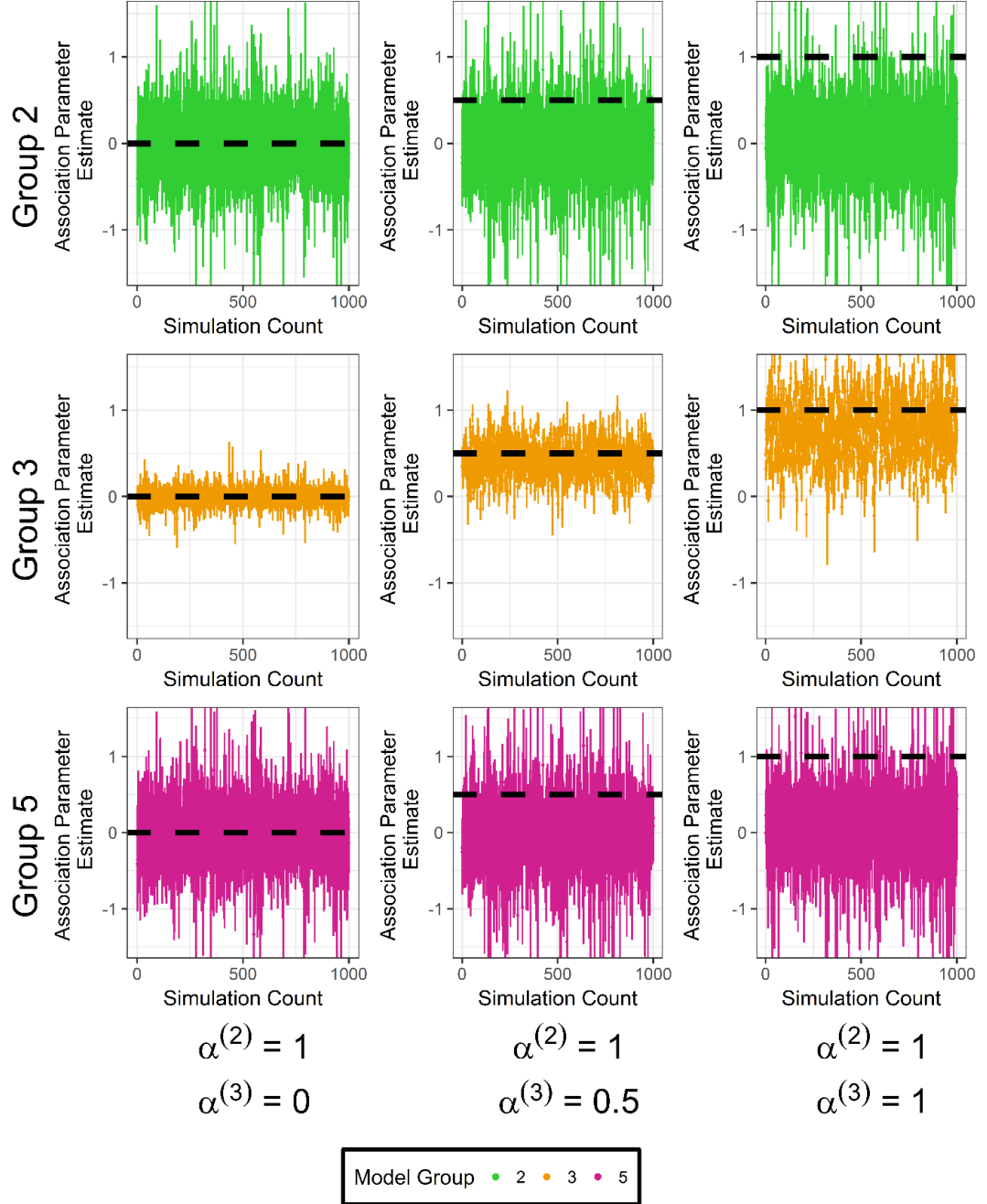


Figure 51: Point estimates and confidence intervals for study level association parameter  $\alpha^{(3)}$  for simulation set 1 investigating varying association parameters (values of association parameters that the data was simulated under are stated under each column). The dashed line indicates the value of  $\alpha^{(3)}$  that the data was simulated under

## Simulation Set 2: Varying number of included studies

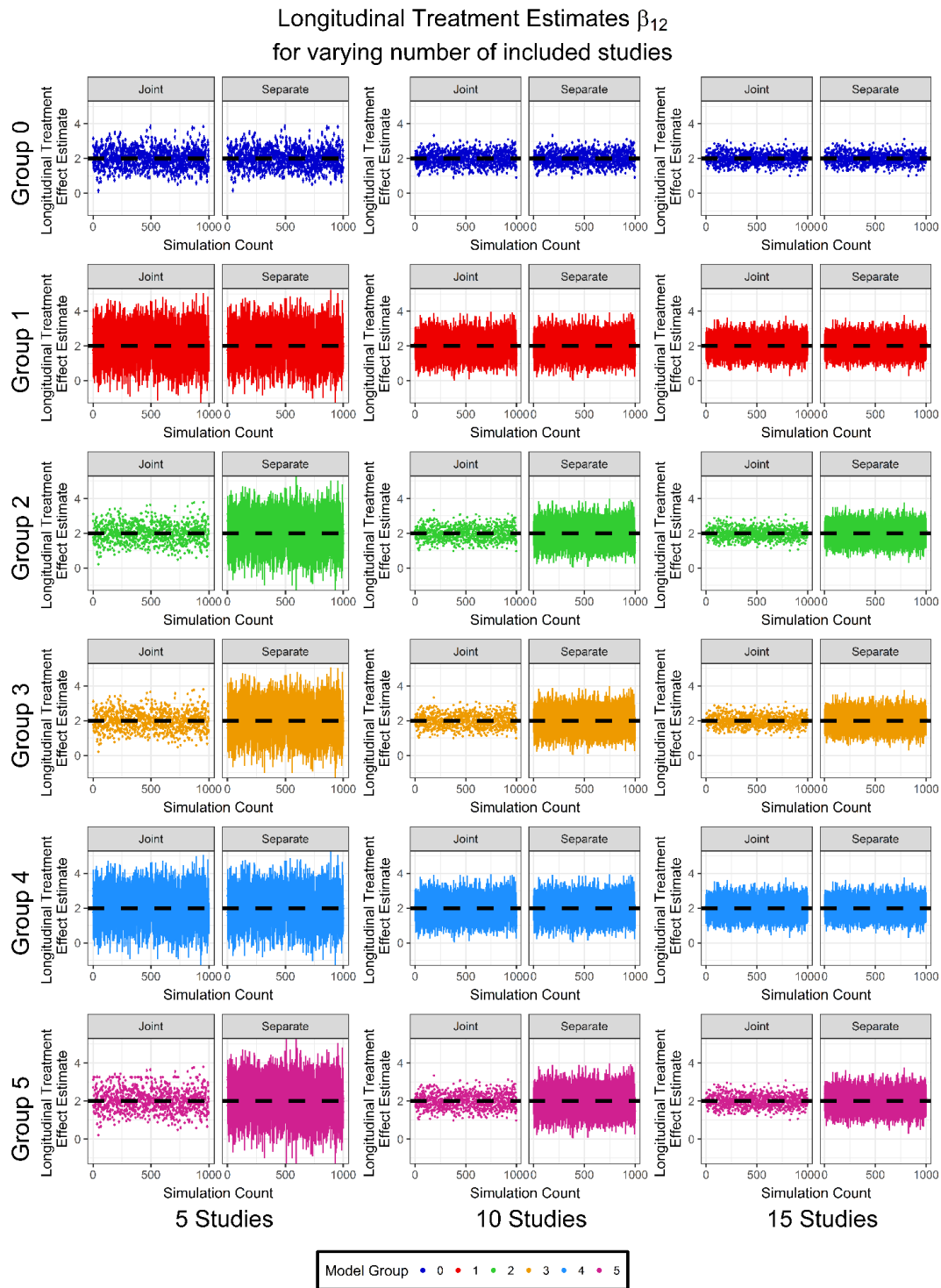


Figure 52: Point estimates and confidence intervals for longitudinal treatment effect ( $\beta_{12}$ ) for simulation set 2 investigating varying number of included studies. The dashed line indicates the value of  $\beta_{12}$  that the data was simulated under

# Time-to-Event Treatment Estimates $\beta_{21}$ for varying number of included studies

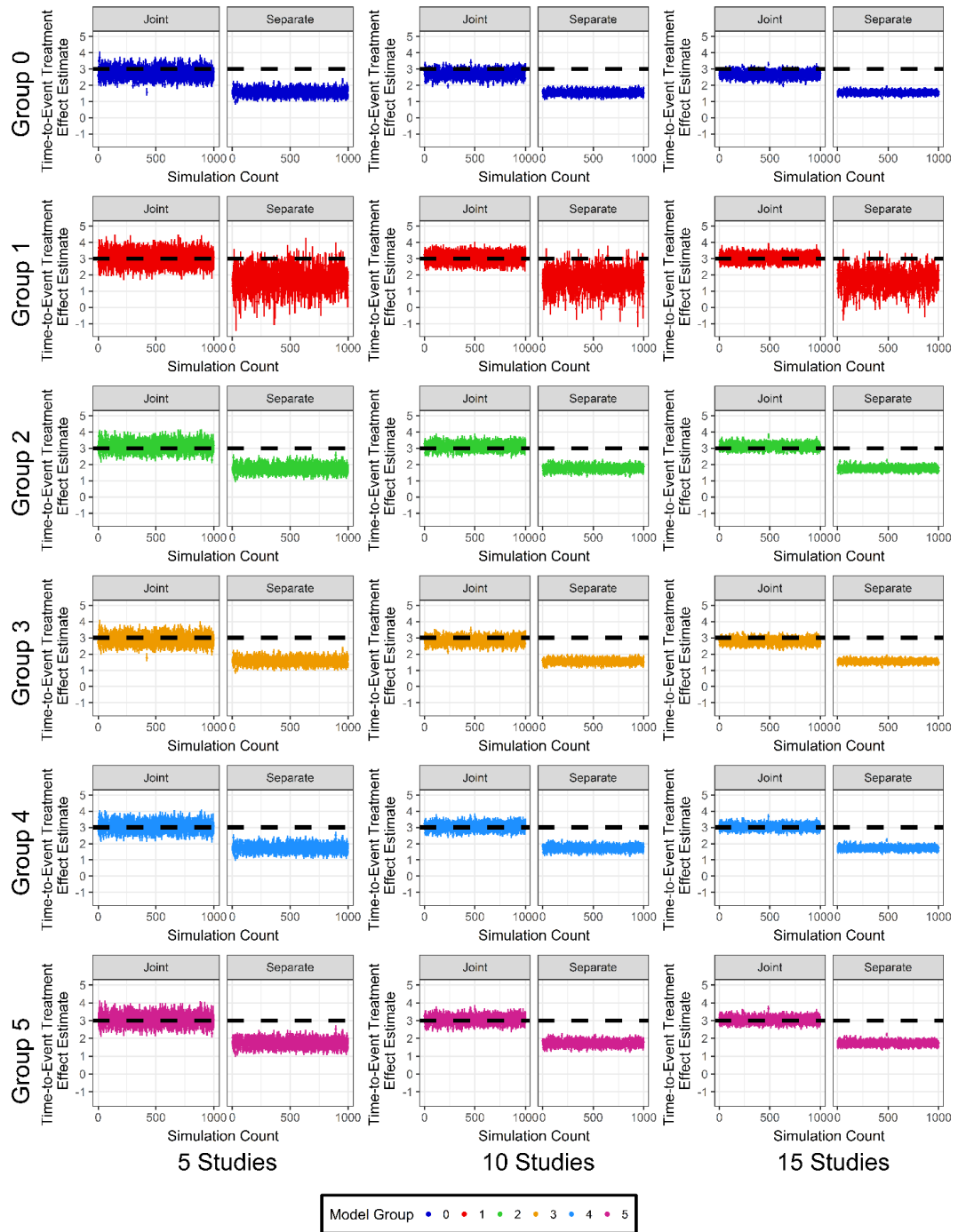


Figure 53: Point estimates and confidence intervals for time-to-event treatment effect ( $\beta_{21}$ ) for simulation set 2 investigating varying number of included studies. The dashed line indicates the value of  $\beta_{21}$  that the data was simulated under

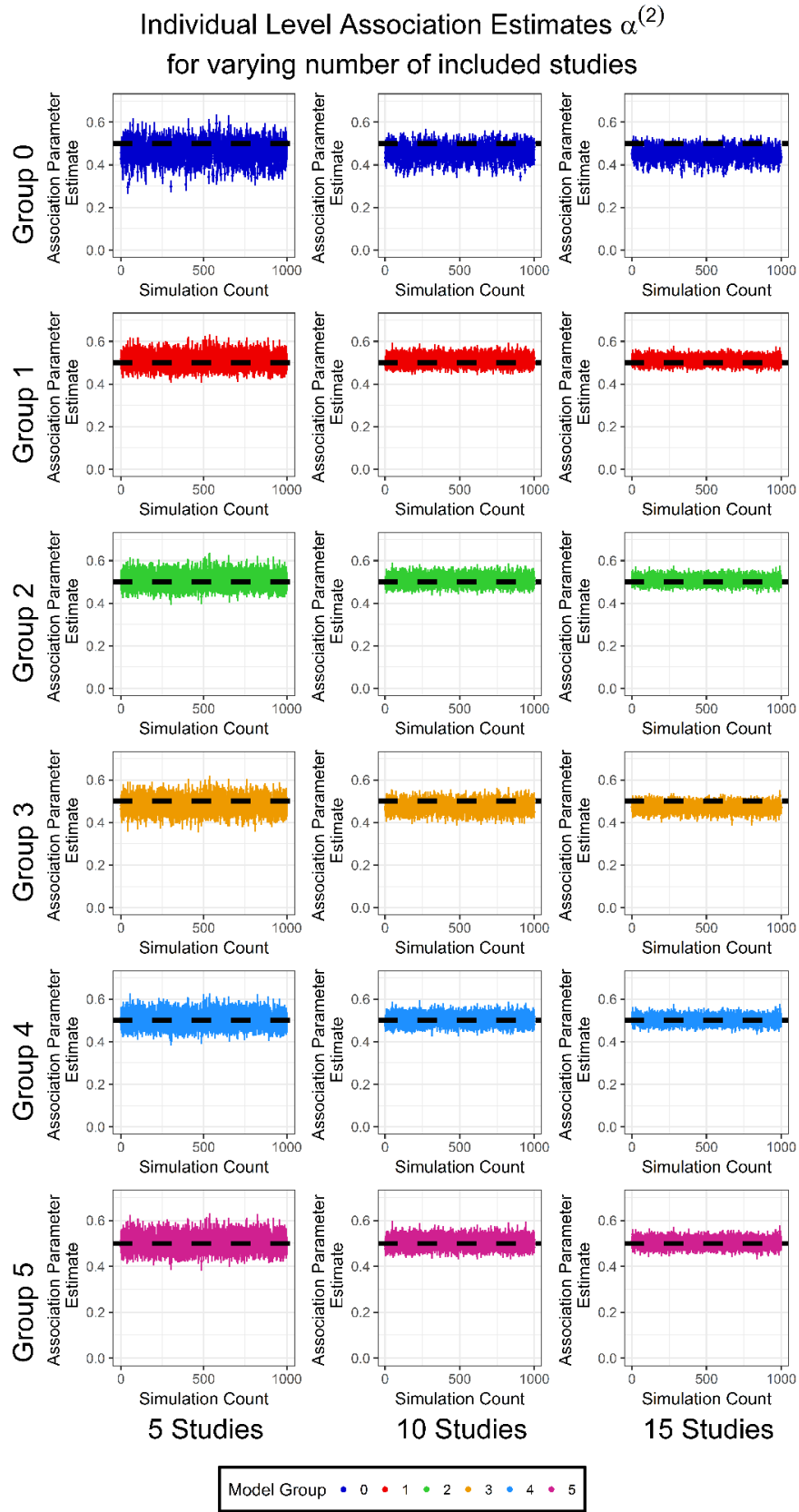


Figure 54: Point estimates and confidence intervals for individual level association ( $\alpha^{(2)}$ ) for simulation set 2 investigating varying number of included studies. The dashed line indicates the value of  $\alpha^{(2)}$  that the data was simulated under

## Study Level Association Estimates $\alpha^{(3)}$ for varying number of included studies

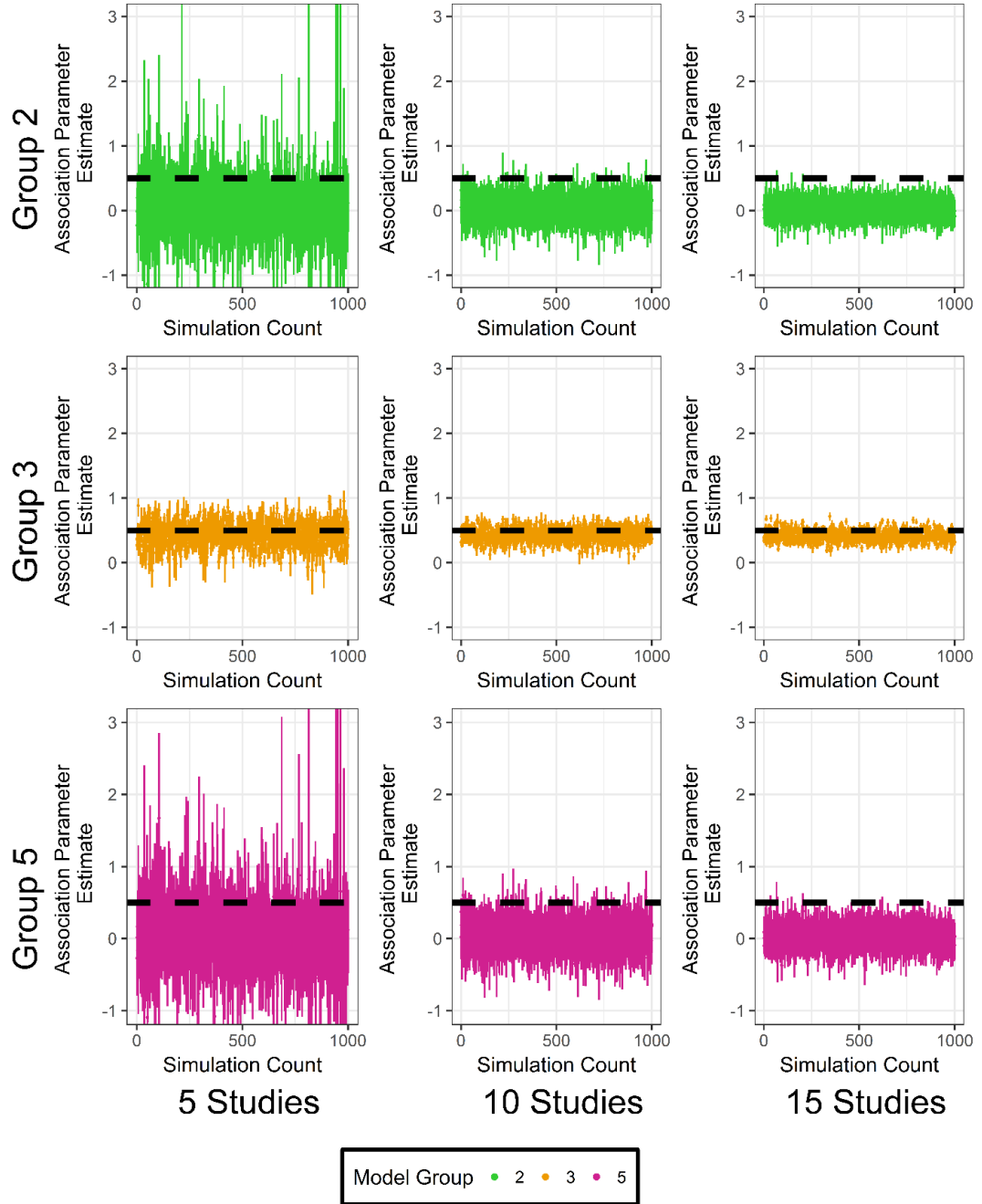


Figure 55: Point estimates and confidence intervals for study level association ( $\alpha^{(3)}$ ) for simulation set 2 investigating varying number of included studies. The dashed line indicates the value of  $\alpha^{(3)}$  that the data was simulated under

### Simulation Set 3: Varying levels of between study heterogeneity

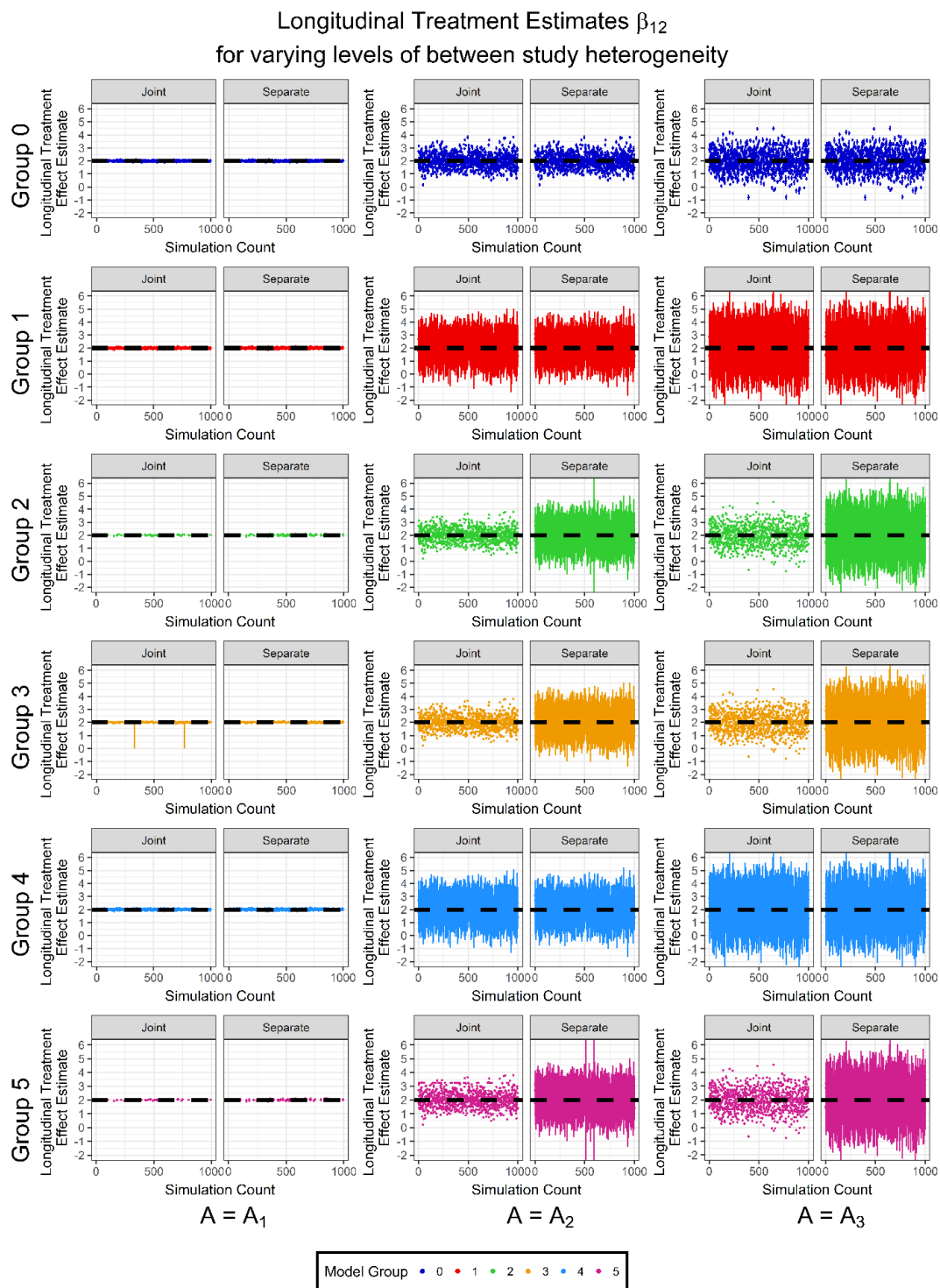


Figure 56: Point estimates and confidence intervals for longitudinal treatment effect ( $\beta_{12}$ ) parameter for simulation set 3 investigating varying between study heterogeneity. The dashed line identifies the value of  $\beta_{12}$  that the data was simulated under



Time-to-Event Treatment Estimates  $\beta_{21}$   
for varying levels of between study heterogeneity

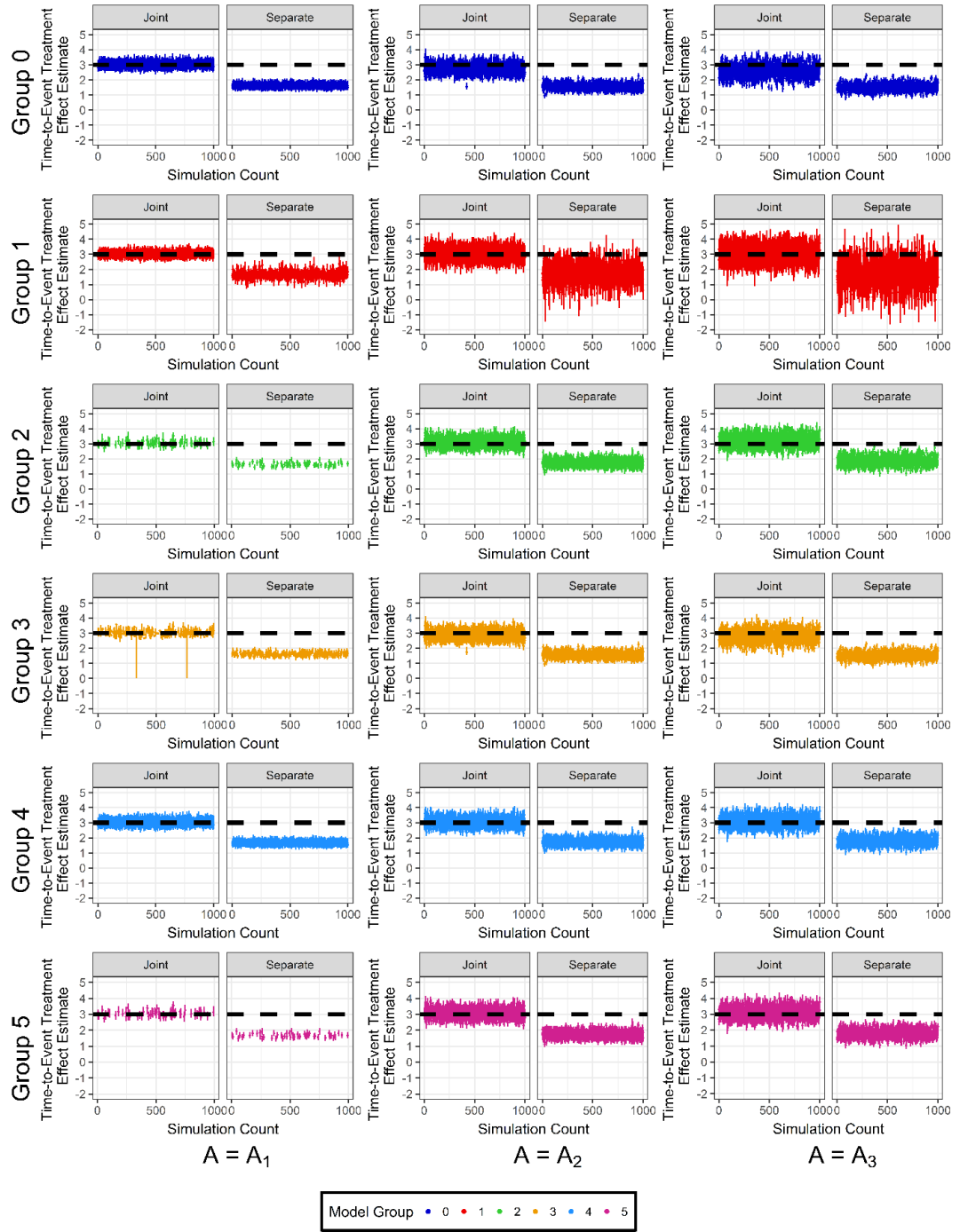


Figure 57: Point estimates and confidence intervals for time-to-event treatment effect ( $\beta_{21}$ ) parameter for simulation set 3 investigating varying between study heterogeneity. The dashed line identifies the value of  $\beta_{21}$  that the data was simulated under.

Individual Level Association Estimates  $\alpha^{(2)}$   
for varying levels of between study heterogeneity

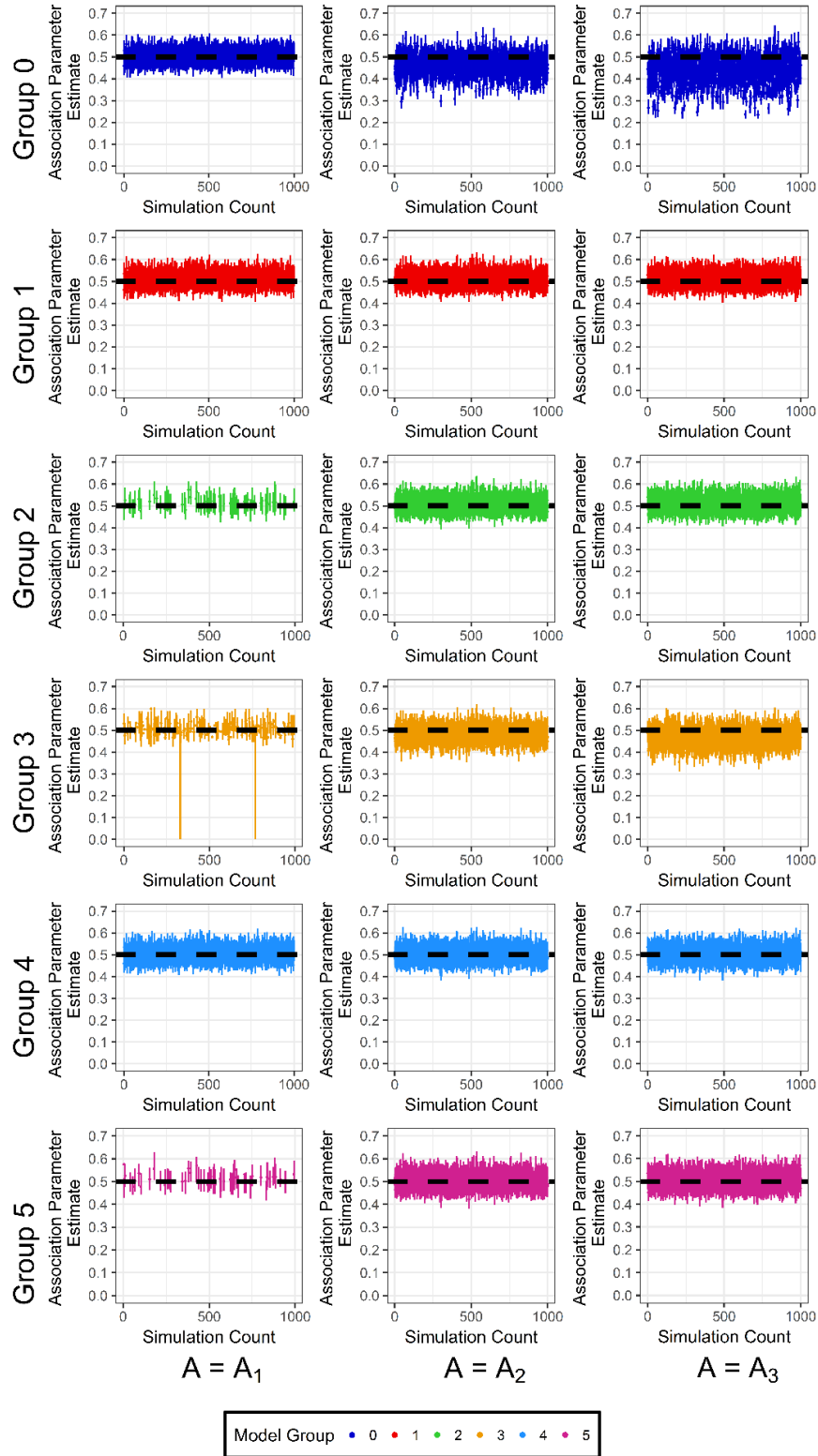


Figure 58: Point estimates and confidence intervals for individual level association parameter ( $\alpha^{(2)}$ ) parameter for simulation set 3 investigating varying between study heterogeneity. The dashed line identifies the value of  $\alpha^{(2)}$  that the data was simulated under.

# Study Level Association Estimates $\alpha^{(3)}$ for varying levels of between study heterogeneity

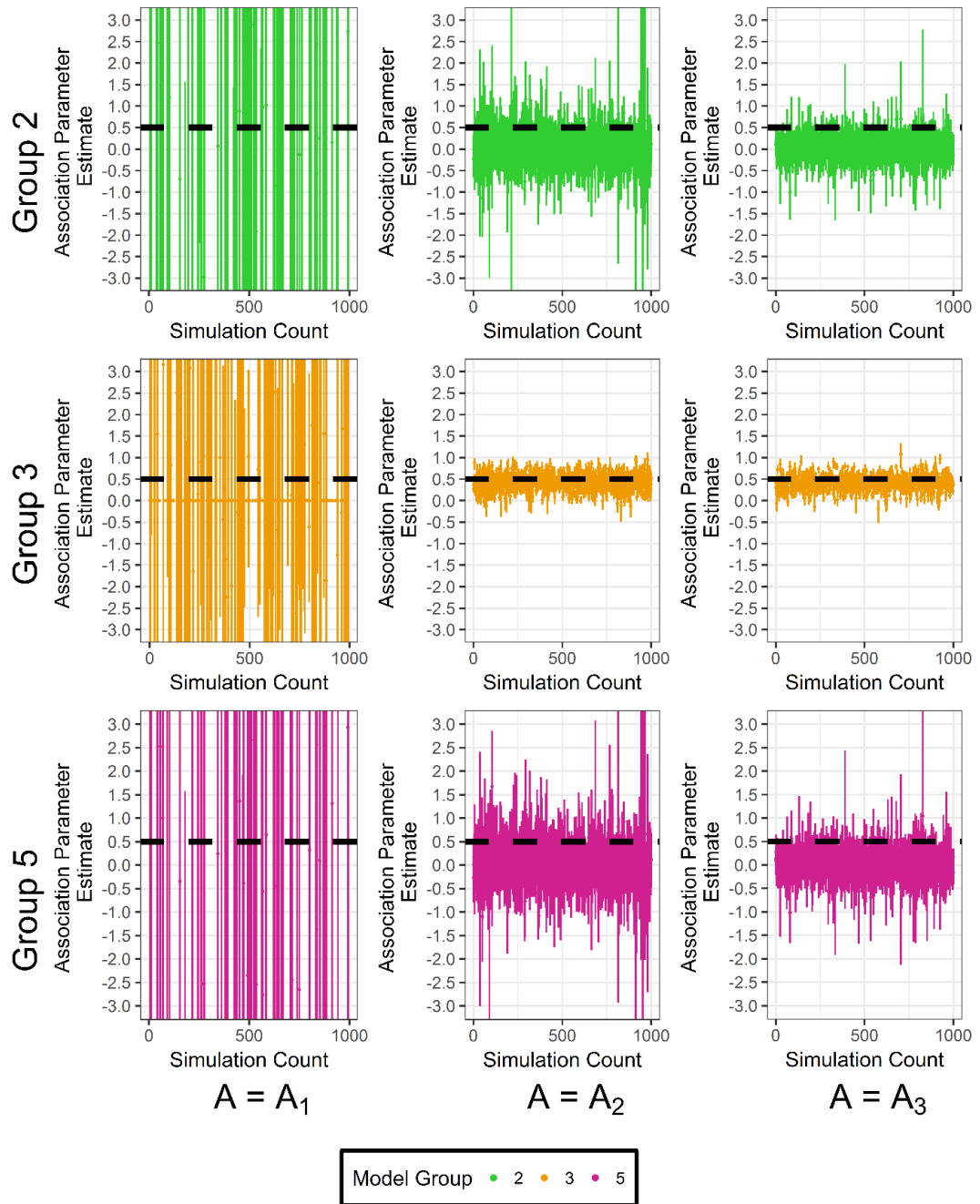


Figure 59: Point estimates and confidence intervals for study level association parameter ( $\alpha^{(3)}$ ) parameter for simulation set 3 investigating varying between study heterogeneity. The dashed line identifies that value of  $\alpha^{(3)}$  that the data was simulated under

## Appendix 5: Association Structure Estimates for One-Stage Analyses of INDANA data

Model Group	COOP	EWPH	MRC1	MRC2	SHEP	STOP	OVERALL
DEATH	0	0.72 (0.44, 1.07) [-0.73, 3.45]	0.3 (0.06, 0.61) [-0.75, 2.05]	-0.22 (-0.48, 0.06) [-1.53, 1.82]	0.27 (0.01, 0.55) [-1.37, 1.95]	-0.02 (-0.24, 0.24) [-1.23, 1.95]	0.68 (0.42, 0.98) [-0.36, 1.89]
	1	-0.03 (-0.14, 0.1) [-0.66, 1.14]	-0.02 (-0.12, 0.1) [-0.51, 0.71]	-0.01 (-0.11, 0.11) [-0.55, 0.87]	-0.03 (-0.14, 0.09) [-0.72, 0.71]	-0.02 (-0.11, 0.09) [-0.5, 0.81]	-0.01 (-0.12, 0.1) [-0.42, 0.44]
	2	-0.03 (-0.14, 0.1) [-0.66, 1.14]	-0.02 (-0.12, 0.1) [-0.5, 0.71]	-0.01 (-0.11, 0.11) [-0.55, 0.86]	-0.03 (-0.14, 0.09) [-0.72, 0.71]	-0.02 (-0.11, 0.09) [-0.5, 0.81]	0 (-0.09, 0.1) [-0.41, 0.44]
	3	0.85 (0.75, 0.96) [0.3, 1.87]	0.05 (-0.04, 0.17) [-0.38, 0.64]	-1.05 (-1.15, -0.93) [-1.6, -0.35]	-0.05 (-0.15, 0.05) [-0.65, 0.6]	-0.63 (-0.72, -0.53) [-1.11, 0.03]	0.74 (0.64, 0.84) [0.33, 1.19]
	4	-0.03 (-0.14, 0.1) [-0.64, 1.11]	-0.02 (-0.12, 0.1) [-0.49, 0.69]	-0.01 (-0.11, 0.11) [-0.53, 0.84]	-0.03 (-0.14, 0.08) [-0.7, 0.69]	-0.02 (-0.1, 0.08) [-0.49, 0.79]	-0.01 (-0.09, 0.09) [-0.41, 0.43]
MI	0	0.61 (0.38, 0.88) [-0.58, 3.13]	-0.18 (-0.39, 0.06) [-1.25, 1.52]	0.23 (0.01, 0.46) [-1.12, 1.76]	-0.01 (-0.19, 0.2) [-1, 1.58]	0.58 (0.36, 0.81) [-0.28, 1.6]	-0.04 (-0.29, 0.25) [-1.25, 3.13]
	1	-0.04 (-0.21, 0.14) [-0.98, 1.84]	-0.01 (-0.17, 0.17) [-0.82, 1.29]	-0.04 (-0.21, 0.13) [-1.08, 1.1]	-0.03 (-0.16, 0.13) [-0.77, 1.2]	-0.01 (-0.14, 0.14) [-0.58, 0.63]	-0.02 (-0.17, 0.15) [-1.08, 1.84]
	2	-0.04 (-0.21, 0.14) [-0.98, 1.83]	-0.14 (-0.3, 0.04) [-0.91, 1.11]	-0.01 (-0.18, 0.16) [-1.05, 1.13]	-0.12 (-0.26, 0.03) [-0.9, 1.13]	0.2 (0.06, 0.36) [-0.4, 0.9]	-0.1 (-0.26, 0.09) [-1.05, 1.83]
	3	0.54 (0.37, 0.71) [-0.36, 2.35]	-0.67 (-0.83, -0.5) [-1.5, 0.62]	-0.05 (-0.21, 0.12) [-1.03, 1.05]	-0.41 (-0.54, -0.27) [-1.13, 0.72]	0.48 (0.35, 0.63) [-0.12, 1.14]	-0.51 (-0.73, -0.19) [-1.5, 2.35]
	4	-0.04 (-0.21, 0.14) [-0.96, 1.81]	-0.01 (-0.16, 0.16) [-0.81, 1.27]	-0.04 (-0.21, 0.13) [-1.06, 1.08]	-0.03 (-0.16, 0.13) [-0.76, 1.17]	-0.01 (-0.14, 0.13) [-0.56, 0.62]	-0.02 (-0.17, 0.15) [-1.06, 1.81]
ST	0	-0.04 (-0.21, 0.14) [-0.97, 1.81]	-0.14 (-0.3, 0.04) [-0.9, 1.1]	-0.01 (-0.18, 0.16) [-1.04, 1.12]	-0.12 (-0.25, 0.03) [-0.89, 1.11]	0.2 (0.06, 0.35) [-0.39, 0.89]	-0.1 (-0.26, 0.09) [-1.04, 1.81]
	1	1.01 (0.6, 1.47) [-2.18, 4.08]	-0.29 (-0.65, 0.1) [-2.11, 2.51]	0.39 (0.03, 0.77) [-1.85, 2.85]	-0.01 (-0.31, 0.33) [-1.77, 2.54]	0.95 (0.59, 1.35) [-0.6, 2.62]	-0.06 (-0.48, 0.42) [-2.18, 4.08]
	2	-0.07 (-0.37, 0.25) [-2.69, 2.24]	-0.02 (-0.29, 0.29) [-1.43, 2.25]	-0.07 (-0.36, 0.23) [-1.93, 2.02]	-0.04 (-0.27, 0.22) [-1.45, 1.98]	-0.02 (-0.24, 0.24) [-1.25, 1.12]	-0.03 (-0.3, 0.27) [-2.69, 2.25]
	3	-0.09 (-0.39, 0.23) [-2.7, 2.22]	-0.29 (-0.57, 0.03) [-1.79, 2.07]	-0.01 (-0.3, 0.29) [-1.84, 2.1]	-0.23 (-0.47, 0.02) [-1.59, 1.84]	0.43 (0.18, 0.72) [-0.67, 1.7]	-0.2 (-0.5, 0.13) [-2.7, 2.22]
	4	0.88 (0.61, 1.18) [-1.44, 2.97]	-1.11 (-1.37, -0.83) [-2.43, 0.84]	-0.07 (-0.33, 0.2) [-1.74, 1.78]	-0.68 (-0.89, -0.44) [-2, 1.1]	0.81 (0.58, 1.04) [-0.4, 1.87]	-0.84 (-1.21, -0.32) [-2.43, 2.97]
ST	0	-0.07 (-0.37, 0.25) [-2.69, 2.23]	-0.02 (-0.29, 0.29) [-1.43, 2.25]	-0.07 (-0.36, 0.23) [-1.92, 2.02]	-0.04 (-0.27, 0.22) [-1.45, 1.98]	-0.02 (-0.24, 0.24) [-1.25, 1.12]	-0.03 (-0.3, 0.26) [-2.69, 2.25]
	5	-0.09 (-0.39, 0.23) [-2.7, 2.21]	-0.28 (-0.57, 0.03) [-1.79, 2.06]	-0.01 (-0.3, 0.29) [-1.84, 2.1]	-0.23 (-0.47, 0.03) [-1.58, 1.84]	0.43 (0.18, 0.71) [-0.67, 1.7]	-0.2 (-0.49, 0.14) [-2.7, 2.21]

Table 29: Table displaying the median (lower quartile, upper quartile) [minimum observed value, maximum observed value] for each study and overall for the value of  $W_{2ki}(t)$  for each model group fitted to SBP and time to death, SBP and time to Myocardial Infarction (MI), and SBP and time to stroke (ST).

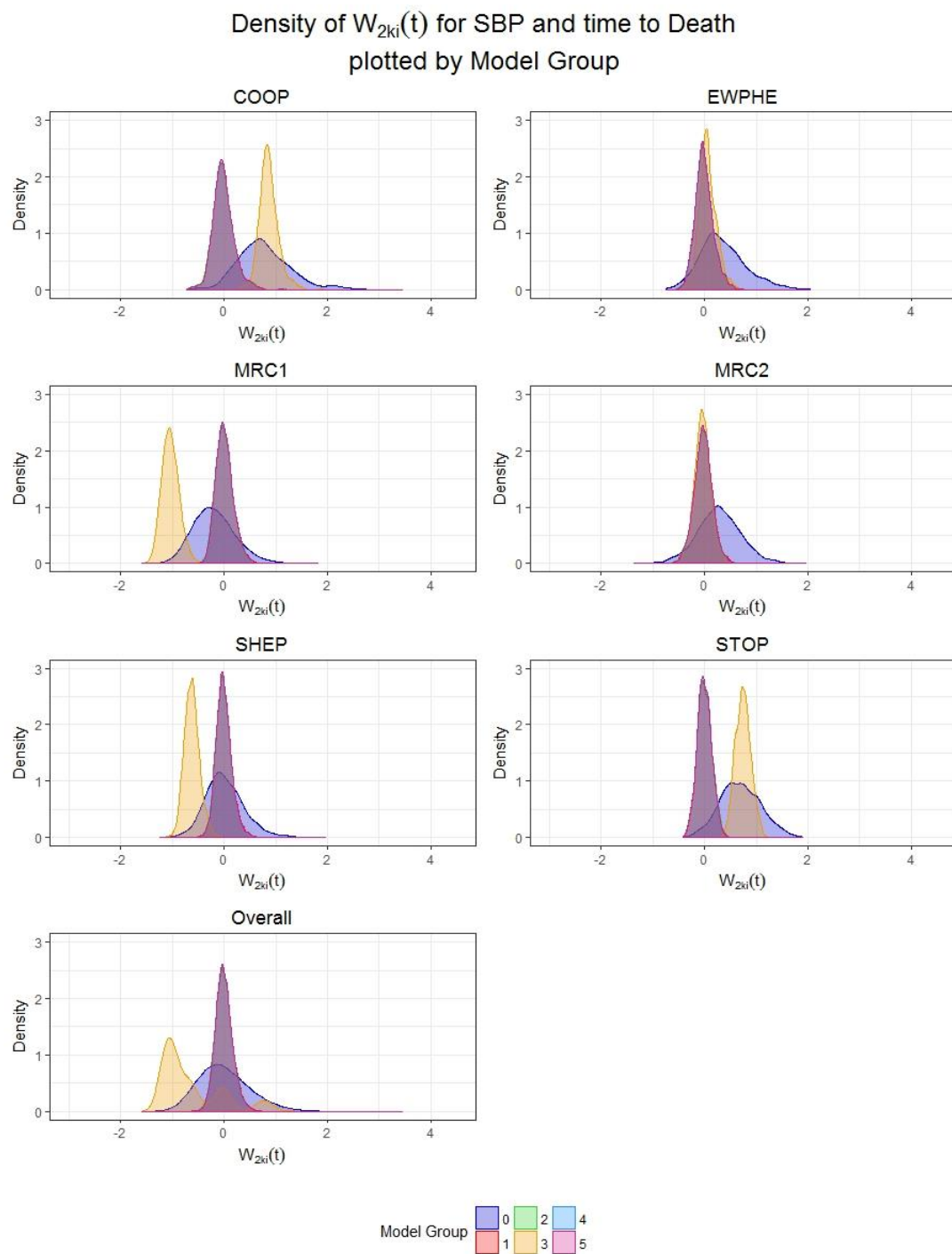


Figure 60: Plot of densities of estimated  $W_{2ki}(t)$  for each fitted model group, for analysis of SBP and time to death, for each study separately and overall for the meta-dataset

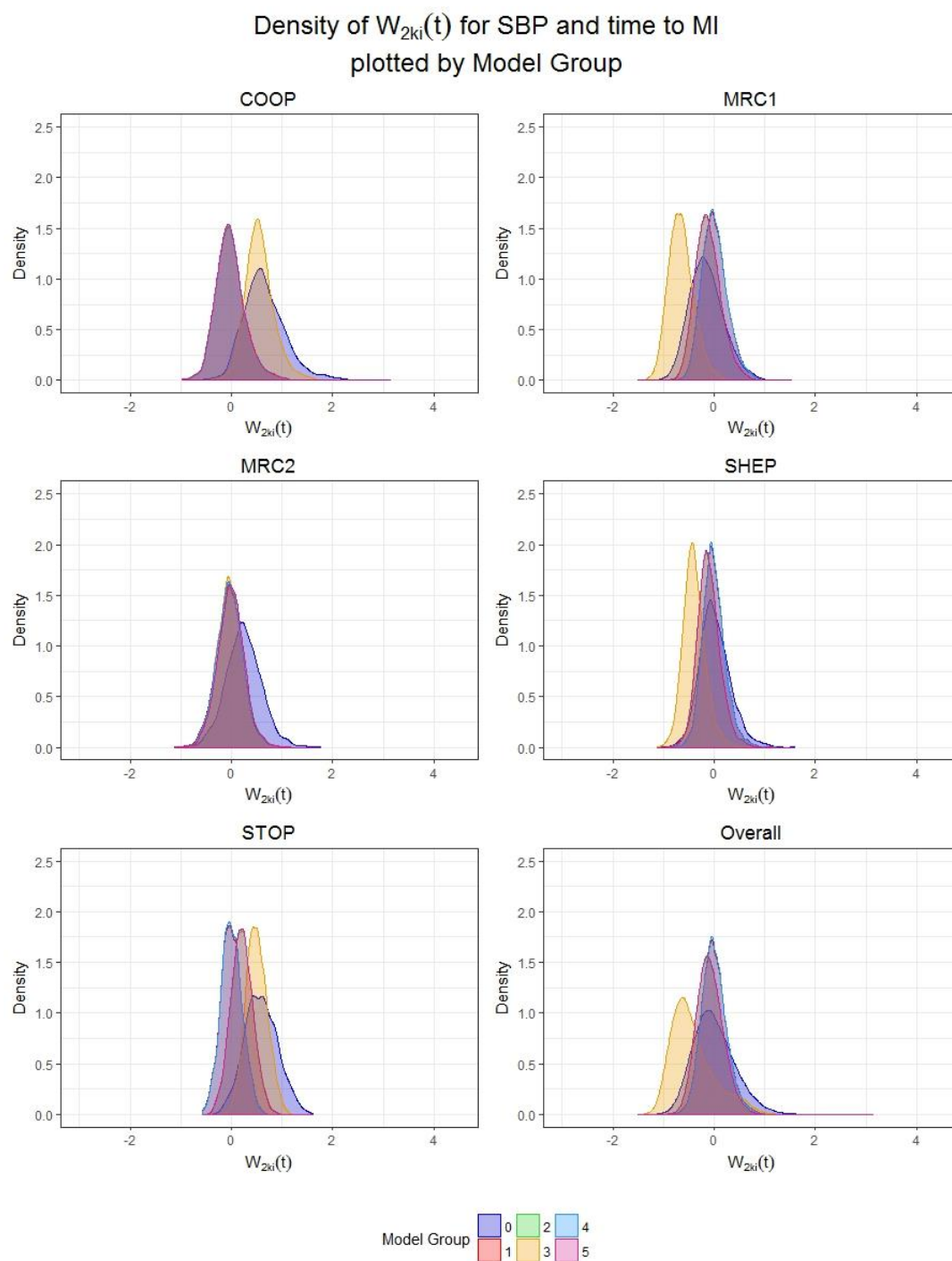


Figure 61: Plot of densities of estimated  $W_{2ki}(t)$  for each fitted model group, for analysis of SBP and time to Myocardial Infarction (MI), for each study separately and overall for the meta-dataset

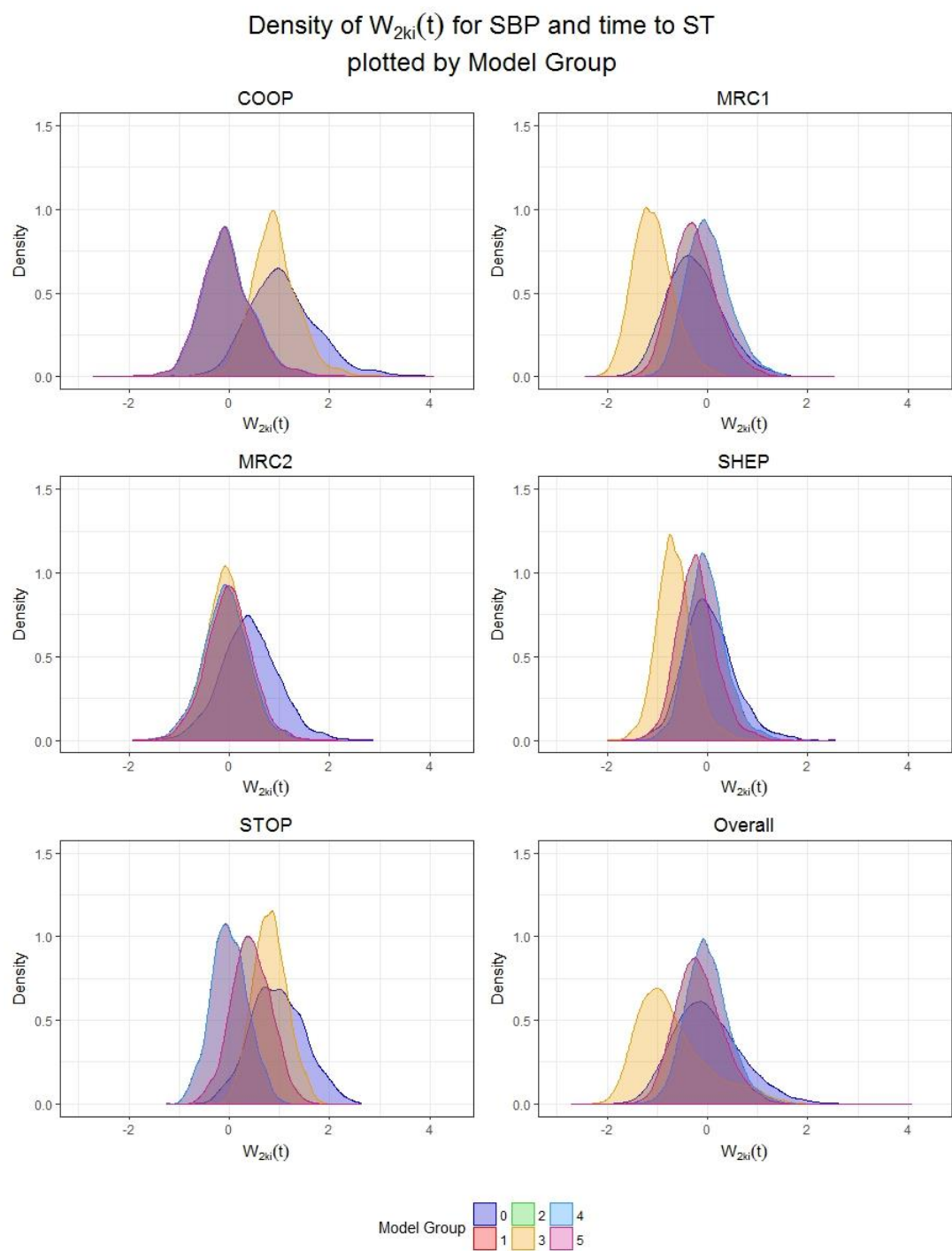


Figure 62: Plot of densities of estimated  $W_{2ki}(t)$  for each fitted model group, for analysis of SBP and time to stroke (ST), for each study separately and overall for the meta-dataset